National Society for the Study of Education

TWENTY-SEVENTH YEARBOOK, PART II

Nature and Nurture
Their Influence upon Achievement

THE

TWENTY-SEVENTH YEARBOOK

OF THE

NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

NATURE AND NURTURE

PART II

THEIR INFLUENCE UPON ACHIEVEMENT

Prepared by the Society's Committee

W. C. Bagley, B. T. Baldwin, C. C. Brigham, F. N. Freeman, Rudolf Pintner, G. M. Whipple, and L. M. Terman (*Chairman*)

Assisted by

L. D. Anderson, Barbara S. Burks, M. C. Barlow, Margaret V. Cobb, Katharine M. Denworth, June E. Downey, P. R. Farnsworth, H. L. Hartshorne, J. D. Heilman, Leta S. Hollingworth, T. C. Holy, Lonzo Jones, M. A. May, W. A. McCall, F. P. Obrien, Joseph Peterson, G. M. Ruch, Howard Taylor, and M. J. Van Wagenen

---

Edited by

GUY MONTROSE WHIPPLE

THIS YEARBOOK WILL BE DISCUSSED AT THE BOSTON MEETING OF THE NATIONAL SOCIETY, FEBRUARY 25 AND 28, 1928, 8:00 P. M.

PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS
1928

OFFICERS OF THE SOCIETY

for 1927-8

Board of Directors

(Term of office expires March 1st of the year indicated)

MEMBERSHIP
OF THE SOCIETY'S YEARBOOK COMMITTEE
ON
THE POSSIBILITIES AND LIMITATIONS OF TRAINING

William C. Bagley, Professor of Education, Teachers College, Columbia University, New York City, New York

Bird T. Baldwin, Research Professor in Educational Psychology, State University of Iowa, Iowa City, Iowa

Carl C. Brigham, Associate Professor of Psychology, Princeton University, Princeton, New Jersey

Frank N. Freeman, Professor of Educational Psychology, The University of Chicago, Chicago, Illinois

Rudolf Pintner, Professor of Educational Psychology, Teachers College, Columbia University, New York City, New York

Lewis M. Terman *(Chairman)*, Professor of Psychology, Stanford University, Palo Alto, California

Guy M. Whipple, Director of the National Intelligence Tests, and Editor of the Society's Yearbooks, Danvers, Massachusetts

TABLE OF CONTENTS

SECTION II. Factors Influencing Special Traits

SECTION III. Summaries of Related Studies

EDITOR'S PREFACE

The Yearbook on Nature and Nurture has grown out of a sug gestion made by Dr. L. M. Terman in December, 1923, that th Society might profitably initiate some research on the limitation of educability, might help to answer the question whether edu cational effort can or can not make bright children out of du ones. The Executive Committee of the Society, at a meeting hel at Detroit, January, 1924, acted favorably upon the suggestio and requested Dr. Terman, in collaboration with Dr. Bagley, t select a suitable committee to undertake the work. In April, 1924 the plans were formally endorsed and a Committee appointed t produce a Yearbook on the "Possibilities and Limitations of Train ing." Further development of the plans has been set before th Society already (for example, in the 24th Yearbook, Part II, pp 386-8, and in the 25th Yearbook, Part II, pp. 254-6). The chang of title of the Yearbook itself does not, of course, involve any essen tial change in the subject matter dealt with by the Committee dur ing the four years the work has been under way.

The Board of Directors appropriated for the expenses of thi Yearbook Committee $700 in April, 1924, $600 in October, 1925 and a special fund of $500 in February, 1927. These sums, how ever, represent only a fraction of the money expended in procurin the data that follow. Four of the major contributions—the Chi cago and Stanford investigations of foster children (Chapters IX and X of Part I), the Hollingworth-Cobb investigation (Chapte I of Part II), and the Heilman investigation (Chapter II of Par II)—have been facilitated by grants from the Commonwealt Fund, from Mr. Max Rosenberg, from Stanford University, fro the Institute of Educational Research of Teachers College, from th Carnegie Corporation of New York, from the Colorado State Teach ers College, and from the Denver Public Schools. The Societ expresses its appreciation of these contributions; explicit acknowl edgment of them is made elsewhere in these volumes. Many othe organizations and institutions have also assisted us, especiall through permitting their representatives to devote their time an energy to the gathering of data and preparation of contributions If these contributions and the personal efforts of the numerou

persons concerned could be computed also in terms of dollars and cents, the investment of capital in the production of the Twenty-Seventh Yearbook would indeed be impressive. The book presents a splendid example of the kind of coöperative research which the Board of Directors has aimed to develop through its policy of establishing committees of experts, supplying them directly or indirectly with adequate funds, over a period of several years, and providing a suitable medium for the publication and dissemination of their reports.

The fundamental importance of the topic under discussion in this Yearbook is clearly set forth in Dr. Terman's introductory chapter. Nothing more need be said on that point, but it may be well here to warn the reader that the problem is not only important, but also difficult. It has been impossible in this volume to preserve the relatively simple style of presentation that has characterized many of our yearbooks which were designed to be directly helpful to the classroom teacher. It has been necessary in the present series of studies to employ intricate methods of approach and complex statistical methods of treatment. That could not be avoided, but the committee has sought to assist the lay reader by prefacing each chapter with a brief orientation, explaining the purpose, method, and outcome of the study in question and its relation to the general problem of the Yearbook.

Those who think of the nature-nurture issue as a fruitful field for rhetorical debate will be disappointed in the pages that follow. No one has tried to be controversial. Every contributor has sought to be straightforward and objective. The aim in this Yearbook has not been to compare opinions (however helpful that may have been in dealing with other educational issues), but to determine facts.

With these explanations in mind, the editor trusts the members of the Society will approve of the content and style of the Twenty-Seventh Yearbook, despite its technical vein and the somewhat abstruse character of many of the contributions.

G. M. Whipple

GENERAL INTRODUCTION TO THE 1928 YEARBOOK

LEWIS M. TERMAN
Stanford University
Chairman of the Yearbook Committee

Background of the 1928 Yearbook. For several years a lively controversy has waged among school people and psychologists with reference to the causes and significance of the large individual differences which intelligence and achievement tests have disclosed. The issues involved are obviously of great consequence for educational theory and practice, since they are part and parcel of the age-old problem as to the relative influence of nature and nurture upon human destinies. If the differences found are due in the main to controllable factors of environment and training, then, theoretically, at least, they can be wiped out by appropriate educational procedures—procedures which it would then become our duty to provide. On the other hand, if they are due primarily to differences in original endowment, then the duty of the school is clearly to provide for differentiated training which will take these native differences into account.

Now any proposal to apply in practice the doctrine of the biological inequality of human beings, whether in politics or education, impinges upon some of our deepest-lying prejudices and challenges our fundamental philosophies of life. It was only natural, therefore, that all too frequently the controversy should have descended from the level of calm discussion of the pertinent scientific evidence to that of acrimonious asseveration and emotional appeal. Rightly or wrongly, some have felt that educational democracy is at stake, and any threat, fancied or real, against the democratic ideal of public education arouses the same antagonism as a threat to the democratic principle of political control. Catch phrases, like "educational determinism" or "democracy and the I.Q.," have become as charged with emotion as were once such slogans as "states rights," "abolition," or "taxation without representation." The belief in native differences and in the validity of intelligence tests as a method of measuring them is looked upon by some of the environmentalists as belonging to the same category

as the theological doctrine of infant damnation. Unfortunately, the dogma of certain of the environmentalists has too often been answered by the dogma of hereditarians. Too many on both sides have shown more interest in finding and phrasing arguments to support a theory than in investigating the problem by the methods of science.

Not that there has not been ample reason in the past why speculations rather than facts have tinged our thinking upon this question. As a scientific problem, this one is exceedingly young. Until Galton (without proving whether the fact was due to nature or to nurture) showed, when he published his *Hereditary Genius* in 1869, that high ability has a strong tendency to cluster in family lines, no one had ever made an extensive investigation in this field. Although *Hereditary Genius* was the stimulus for a few lesser studies which followed it closely in time, it was not until the beginning of the twentieth century that concerted resources were devoted to unraveling the problem. Even then the early attempts were of necessity tentative and incomplete, partly because the problem itself was as yet merely sensed rather than clearly formulated, partly because a large amount of preliminary exploratory work was essential before the lines of attack most likely to be fruitful and crucial could be appreciated, and partly because refined methods of measuring intellectual abilities were not yet available.

Possibly one misconception (now fortunately clarified by recent contributors to the nature-nurture field) has clouded the pertinent issues as much as any single other. This is the notion that a 'nature' theory implies a type of 'glass bottle' mental development which is quite independent of any stimulation from environment. No idea could be more misleading, and no phenomenon more impossible. The theoretical as well as the practical nature-nurture problem that faces us deals with human beings as they are—not as an experimenter might rear them in a nurture-proof laboratory. We are interested in the child or adult as he comes to us—with his unique complex of ancestry, associates, home training, schooling, and physical and moral attributes. We are interested, not in finding out how he would have developed if he had had no environment at all; rather we wish to discover whether or not he can be made a more intelligent individual om a more learned one by improving

the conditions of his *milieu* within the limits found in reasonably good social communities. More generally, we wish to find the relative potency of all types of human environment to add to, or to detract from, human endowment, and to know the limits placed upon achievement by endowment.

Inception of the Present Yearbook Project. The situation set forth in the preceding paragraphs was discussed in the February, 1924, meeting of the Board of Directors of the National Society for the Study of Education, and following this discussion an agreement was reached to devote an early yearbook to the topic, "The Possibilities and Limitations of Training." Shortly afterward I was invited by the Board of Directors to form a committee for the preparation of a yearbook devoted to this subject. I agreed to do so on condition that the project be made a research undertaking rather than a spectacular combat of educational gladiators. In view of the large amount of heated discussion that had taken place, it seemed best to devote this yearbook almost entirely to the dispassionate exposition of the results of new investigations bearing on the nature-nurture problem, with only a minimum of space devoted to orientational and interpretative contributions.

A committee was at once formed which was favorable to such a program. The Committee was selected so as to represent all shades of opinion on the issues involved, for it seemed very important to avoid the suspicion of making the Yearbook the special mouthpiece of either side to the controversy. Whatever faults may be found with the outcome of the Committee's work, I think it can be said that all of its members have shown throughout an objective-minded, investigational attitude. The sole concern of the Committee has been to bring together all the new evidence it was possible to secure on either side of the question. This method of approach once agreed upon, one finds that it is possible for persons of the most antagonistic beliefs to coöperate effectively in a program of research.

Activities of the Committee. The Committee has had only two meetings, one in Chicago in October, 1925, the other in Philadelphia in December, 1926. The Chicago meeting was fully attended and occupied an entire day. The Philadelphia meeting was briefer and could be attended only by Baldwin, Brigham, Terman, and Whipple. Fortunately, the Committee chairman (Terman) and

the Yearbook editor (Whipple) were able to hold a series of conferences at Stanford University in May and June, 1927, and to agree upon various matters with reference to the selection and arrangement of material.

Previous to the first meeting of the committee, the general plan of the Yearbook had been determined upon by correspondence. This plan included the following specifications:

1. The Yearbook should be devoted to the nature-nurture problem, with special reference to the relative influence of these factors upon scores in intelligence and achievement tests.

2. The major part of the space should go to the exposition of new investigational results.

3. To this end, effort should be made to secure the coöperation of as many scientific workers as possible in the conduct of researches bearing on various aspects of the problem.

4. Investigations tending to emphasize *either* the *possibilities* or the *limitations* of training were to be equally welcomed.

5. Contributors would be expected to confine themselves primarily to the exposition of factual data and to the brief statement of their conclusions based thereon.

In October, 1925, an appeal for coöperation was mailed to nearly two hundred persons throughout the country, and this was followed by two published appeals, one in the *Journal of Educational Research*, May, 1926, the other in *School and Society*, March, 27, 1926. Systematic personal enquiries were made for the purpose of getting in touch with investigations already under way which might have a bearing on the Yearbook's problem. Other investigators were urged to undertake special researches for the purpose. The two large and meaty volumes here offered are evidence enough of the response secured. The table of contents for the two volumes shows a wide variety in the special phases of the problem which were attacked. The part devoted to factors influencing intelligence-test scores covers such broad and basic topics as "home environment," and such special ones as "coaching." The part dealing with achievement test scores presents critical data upon a number of factors, such as school expenditures and school attendance, whose efficacy to produce educational results we have always taken for granted in the past. The section on special mental traits offers upon

several important problems material which is virtually pioneer work. It would not be easy to estimate the total labor cost which the investigations of these two volumes represent. In all probability it would bulk to more than $100,000 if bought and paid for in the market. Several of the studies would not have been made at this time, had it not been for the stimulus offered by the Yearbook project.

Acknowledgments. In behalf of the Committee I wish to thank all those who have contributed so liberally of their time and effort to make this Yearbook a success. In many cases the contributions have been made at considerable personal sacrifice. The Committee appreciates not only the contributions as such, but also the non-partisan and scientific spirit in which they have been made. There has been little or no rhetorical straining for effect in the exposition or interpretation of results. Few treatments of the nature-nurture problem have consistently maintained so high a level of open-mindedness and intellectual honesty.

Grateful acknowledgment is made to the Commonwealth Fund for a grant of $15,000, divided equally between Chicago and Stanford Universities, in support of the foster children studies. The Stanford grant was supplemented by a contribution of approximately $2,500 from the Thomas Welton Stanford Fund for Psychological Research and by a gift of $1200 from Mr. Max Rosenberg, of San Francisco. The State Teachers College of Greeley, Colorado, and the public school officials of Denver coöperated helpfully in making possible the extensive and laborious investigation by Dr. Heilman. Many other educational institutions and many individuals, not all of whom it is possible to mention by name, have contributed in important ways to the carrying out of the investigations herein reported.

Special acknowledgment is made of the invaluable assistance rendered by Barbara Stoddard Burks, now Mrs. Herman Ramsperger, of Stanford University. Besides carrying out the Stanford foster-children investigation and preparing the masterly summary of nature-nurture literature, Miss Burks served throughout the period of the Yearbook's preparation as my research and secretarial assistant. In this capacity she not only conducted the major part of the routine correspondence, but in addition gave valuable as-

sistance to several contributors, both in planning their researches and in reporting their results. Following her suggestions, several chapters of the Yearbook were largely rewritten and many statistical procedures were modified. Probably no other single individual deserves more credit for whatever merits the Yearbook possesses.

Suggestions for Use of the Yearbook. The amount of real service which these volumes will render to educational theory and practice will depend in no small degree upon the spirit in which their content is approached. In the first place, readers who do not sufficiently appreciate the inherent difficulties and complexities of the problem, particularly the difficulty of securing factual data that are at once crucial and unambiguous, will perhaps be disappointed in the lack of perfect consistency in the results of various contributors. For certainly it must be admitted that no final answer to the nature-nurture question has been attained or even approximated. The most that can justly be claimed is that the bounds of our knowledge have been in some measure extended. Anyone who has reasonable regard for the fundamental importance of the problem will be grateful for any advance in our knowledge, however slight, which the contributions of this Yearbook have accomplished. It is conceivable that the elusive nature of the problem is such as to preclude for a long time to come, if not forever, a complete and final solution. One needs to cultivate patience and the faculty of suspending judgment.

Secondly, in his study of the investigations reported the reader is urged to divest himself, in so far as it is humanly possible, of his emotional bias and predilections. Because of unconscious factors that are so likely to play a rôle, one can not always be sure whether one has or has not attained to a state of complete intellectual detachment, but the effort should nevertheless be made. How absurd, after all, is the attitude of mind which would force truth into any kind of preconceived mold, when only truth that is genuine and undistorted can give us the control over human nature and human destinies that our various institutions, including the school, are intended to exercise! Surely, anyone who is imbued at all with the spirit of modern science should be less con-

cerned about *what* is true than about knowing what *is* true. In order to counteract the inevitable human tendency to bias, the reader will do well to make a conscientious effort to find merit in those investigations which lend least support to the beliefs he has previously been inclined to favor. It is so easy, as Darwin long ago warned us, to forget the evidence that does not happen to agree with our theories!

Finally, I wish to express the profound hope of the Committee that numerous readers of the Yearbook will be stimulated by the very shortcomings and imperfections of the investigations reported to undertake researches which will be better planned, better controlled, more extensive, and therefore more conclusive in their results. It is by this standard, ultimately, that the success of our efforts must be judged.

PREFATORY NOTE TO CHAPTER I

This is the first of the series of contributions assembled in Part II of the Yearbook in which the attempt is made to measure the relative effect of heredity and environment upon achievement.

The investigation reported by Mrs. Hollingworth and Miss Cobb is unique in one respect at least. Ordinarily, when seeking to measure the possible effect of superior intelligence upon educational achievement, a comparison is made between the school achievements of pupils of average intelligence and those of pupils of superior intelligence or between the school achievements of pupils of average intelligence and those of pupils of inferior intelligence or perhaps between the school achievements of pupils of superior and of inferior intelligence. In any such comparison there enters a serious disturbing factor—that of the home conditions or the cultural surroundings or the educational opportunities of the two groups in question; these environmental conditions are usually more or less markedly different and they must affect, to an extent difficult to compute, the actual school achievements of the two groups.

In the present investigation the two groups under comparison both possess markedly superior intelligence; the homes from which they come appear, from careful cultural rating, to be indistinguishable; moreover, the actual school opportunities have been as carefully equalized as human ingenuity could suggest. Under these conditions, there nevertheless appears a series of divergences in educational achievement, ranging from none at all in some types of school work to marked divergence in others. These divergences are presented in a series of tables and graphs and lead to the general conclusion that the possessor of an intelligence quotient of 165 will distinctly surpass the possessor of an intelligence quotient of 145 in certain, relatively complex, lines of educational achievement, even though the home and school surroundings of the two children are equal, the one to the other, in the various factors that may favor school progress.

CHAPTER I

CHILDREN CLUSTERING AT 165 I.Q. AND CHILDREN CLUSTERING AT 146 I.Q. COMPARED FOR THREE YEARS IN ACHIEVEMENT[1]

LETA S. HOLLINGWORTH
AND
MARGARET V. COBB
Teachers College, Columbia University, New York City

The investigation here reported was formulated in 1922 to discover whether children who, by available tests of intelligence, cluster around a mean at + 8 P.E. will outstrip in achievement over a period of years those clustering at + 6 P.E., provided important factors other than I.Q. are equalized for the groups under comparison.

It has been previously established with reasonable certainty that children who *test* by available methods at low degrees of intellect cannot learn as well as children generally can; and that those who *test* at high degrees of intellect learn much more rapidly and adequately than children in general do. These results have convinced many that innate educability is tested by these methods, and that it is now possible by such means to classify children for differentiated education, justly and humanely. Others have, however, remained unconvinced, pointing to the positive correlation between degree of intellect in pupils on the one hand, and quality of parents, home, and school on the other. Since children who are dull by test and who learn poorly come usually from 'unprivileged' homes, it is inferred that not *educability* but educational *opportunity* is the basis of both intelligence quotient and curve of learn-

[1] This report is rendered as part of the work of a Joint Committee, for three years in charge of special opportunity classes for gifted children at Public School 165, Manhattan. The members of this Committee were Mr. Jacob Theobald and Miss Jane E. Monahan, of Public School 165, Miss Margaret V. Cobb, Dr. Grace A. Taylor, and Professor Leta S. Hollingworth, of Teachers College, Columbia University. The work was carried on with the advice of District Superintendent John E. Wade, and in coöperation with the Division of Educational Psychology, of the Institute of Educational Research, at Teachers College. The work was supported by funds granted through the Institute by the Carnegie Corporation of New York.

ing. (Such reasoning would no doubt also explain the fact that children who 'test high' are accelerated in school status, by saying that they are bright because they are accelerated, rather than by saying that they are accelerated because they are bright.)

Selection of Subjects

In order to avoid the ambiguities arising when stupid children (who usually have inferior social and educational environment) are compared in learning with bright children (who usually have excellent social and educational environment), we here present two groups, both testing in the highest one half of one percent of the total population for intelligence as shown by test. From long acquaintanceship with the social and educational backgrounds of young children testing in the top one-half of one percent, we believed that children in the highest one quarter of one per cent differ little if at all from those in the next to the highest one quarter of one per cent in such matters as excellence of home environment (insofar as this can be measured in material terms) and of educational opportunity. Yet so wide is the *range of intellect* included in the *highest one-half of one per cent*, that it is possible in a school population of a million children, such as is found in the public schools of New York City, to collect two groups of pupils, *both* falling within this designated percentile status, yet separated from each other by a distance of about 2 P.E., or about 20 points of I.Q. In our selection of children we aimed at a mean of 145 I.Q. for one group, called our "Lower Group," and of 165 I.Q. for the other, called our "Higher Group." In matching the children we actually achieved a mean of 146 I.Q. for the one and of 165 I.Q. for the other. These compared groups were formed and finally selected in the following manner.

Our Committee, to which reference has been made, was organized in the spring of 1922 to select and to conduct for three years, a special class of young children all testing at or above 150 I.Q. In searching for these children it was inevitable that a large percentage of promising candidates should be found to test *below 150, but above 135 I.Q.* A group of such children was thus incidentally identified; and the suggestion arose to form a special class for them also, conducted with conditions and opportunities identical (except

as stated later in this discussion) with those planned for the group above 150 I.Q., with the purpose of ascertaining whether or not, among children all so highly intelligent, differences of scholastic achievement would appear in accordance with differences in I.Q. as determined by available methods. It seemed to us that groups selected as these were selected, would not differ from each other in achievement, unless *intellect as measured* is the chief determiner of such a difference.

During the three years of experimentation in the special opportunity classes thus initiated at Public School 165, Manhattan, the registers of both classes were maintained at slightly above 25 pupils, which is the minimal register administratively permissible. Some of the children dropped out from time to time, and were replaced by others, since a certain amount of 'turnover' is unavoidable through change of residence and similar accidents. In May, 1925, the investigators were able to list 40 children who had been, and continued to be, members of these special classes from the beginning of the experiment to its close, and who could be paired in the manner determined upon. Twenty of these tested from 152 to 183 I.Q. (Stanford-Binet). Each of these 20 children was then paired with another child of the same age who tested from 1 P.E. to 4 P.E. *lower* in I.Q. The Lower Group thus showed I.Q.'s ranging from 139 to 156. The mean I.Q. of the Higher Group proved to be 164.9; that of the Lower Group, 146.1.

During the three years of experimentation, each child had been tested annually by means of the Stanford Binet. The I.Q.'s upon which matching was based in May, 1925, are in every case the *means* of these three annual tests. The relative status of each child within the two groups is thus a little more reliably determined than would have been the case with one test only,[2] since the correlation of the Stanford Binet with itself on annual trials is not above .95.

Our ideal was to match each pair for birthday age within a month, but in several cases this ideal was not quite realized. The difference between the groups in mean age gives a slight advantage of maturity to our Lower Group (see Table I). This advantage,

---

[2] In at least one case the mean I.Q. is known to be a little too low, because the final I.Q. dropped, on account of the limitations of the scale at the upper extreme. This is the case of the child of Pair 1, whose I.Q. is 183 as the mean of the three tests. His first test at the age of 9 years yielded an I.Q. of 190.

insignificant as it is, is 'spread' over the members of the Lower Group, and does not result from but one instance of a much older child in this group. In obtaining our 20 matched pairs, we exhausted the possibilities among the children available under the criteria adopted for matching.

The children were 100 months old, on the mean, when selected for the classes in 1922. They were about 102 months old, on the mean, when they entered upon special opportunity in September, 1922. They were 135 months old when matched for the comparative groups herein reported, and the mean age was 136 months when the experimental work was concluded at the end of June, 1925.

Sex was not kept as a constant in matching pairs, but the groups had 5 girls and 15 boys in each of them. The two groups were practically identical as regards race; 18 children in each group were Jewish. Of the two non-Jewish members of the Higher Group one was Irish-German and the other, Scotch-Irish. If the two non-Jewish in the Lower Group one was French-English and the other, Scotch-Italian.

TABLE I. SHOWING THE MATCHING BY AGE AND INTELLECT OF THE TWO GROUPS COMPARED IN EXPERIMENTATION

Pair	Higher Group			Lower Group		
	Age in Mos.	First I. Q.	Mean I.Q.	Age in Mos.	First I. Q.	Mean I.Q.
1.....	144	190	183	144	145	149
2.....	128	172	178	130	135	145
3.....	139	175	176	140	145	153
4.....	121	166	173	125	149	156
5.....	142	171	171	143	140	147
6.....	138	171	171	139	141	141
7.....	124	154	171	129	146	151
8.....	131	167	170	132	149	150
9.....	140	162	165	139	139	150
10.....	129	160	163	133	138	144
11.....	136	170	162	135	135	144
12.....	132	162	161	133	135	139
13.....	137	160	160	135	142	146
14.....	143	157	160	146	137	142
15.....	127	163	159	132	143	148
16.....	140	164	158	141	144	139
17.....	142	156	158	147	141	145
18.....	137	141	155	133	138	146
19.....	134	156	152	134	151	143
20.....	131	154	152	134	141	144
Mean	134.75	163.5	164.9	136.20	141.7	146.1

All of the forty children were selected from the public-school population of New York City, many districts of which were represented. The extremes of residence were from the Bronx to Staten Island; from East Fourteenth Street to West Two Hundred and Sixty-Third Street; from Avenue B to Riverside Drive. The grouping by age and by I.Q. is shown in Table I.

After the children were matched, in May, 1925, as shown in Table I, the achievements of the two groups as thus constituted were computed for the preceding three-year period of opportunity. *Achievement was not taken into account at all in the matching,* which was done solely on the basis of age, I.Q., and presence in the classes throughout the whole three years of experimentation.

Equalization of Educational Opportunity

Time Spent in Graded Classes

When the matching was completed, the investigators had before them two groups of 20 gifted children each, separated by a distance of about 2 P.E. on the scale of I.Q. (1 P.E. $= \pm 8$ I.Q., according to Terman's distribution). All of these children had first attended the regular classes of various public schools. It is of interest to us to know whether our two groups had already been differentiated in achievement by promotions received in these grades before entering our special classes.

After our two comparative groups had been selected and paired, we computed the mean school status of each group at the time of segregation in September, 1922. The result shows that the Higher Group then averaged .8 of a term (semester) higher in school grading than did the Lower Group. The overlapping between the two groups in this respect was very marked. Relative grading was but slightly in accordance with mental age as regards individuals. For instance, the child who appears in our Higher Group with an I.Q. of 183 was but one year accelerated in school status. At the age of 9 years he was in the 5th grade, though his M.A. at that time was 17 years, 4 months. Others of much less intelligence had 'skipped' two and even three years, according to the policies and insights of their various principals and teachers.

Time Spent in Special Classes

Both groups were segregated for special opportunity classes at the same early age and were thereafter taught in the same school, under the same administrative auspices. Here they both had access to the same special library, were taught under the same policy of freedom to progress according to capacity, and enjoyed all special and usual privileges and opportunities provided.

Teachers

Each *child* had had from two to six different teachers during the years spent in the graded classes. Each *group* had thus been subject to the influences of so many different teachers previous to segregation as to destroy the influence of any one of them as a constant upon the group.

After segregation, each group was subject to the influence of seven different teachers, six of whom taught every child in both groups, impartially. The seventh teacher, however, was in each instance the classroom teacher, and this teacher did not teach both groups impartially. These classroom teachers were constant influences for three full years, each for her own group, but not for the other to the same extent. It was administratively impossible to shift these two teachers in the middle of the experiment, as from the research point of view alone would have been desirable. All but three of the children in our Higher Group stayed with one teacher throughout the entire three years, and all but seven of those in our Lower Group stayed with the other teacher throughout.

We have carefully scrutinized our data for evidence as to whether or not the influence of these two teachers might be the cause of differences in achievement between the two groups. We believe such influences to have been inappreciable for these four reasons:

(1) Each group was placed with its teacher at the outset solely on the basis of I.Q.'s. And in September, 1922, before being taught by these teachers, the Lower Group was found to have achieved less in graded classes scattered over the city than had the Higher Group. *Subsequent* performance is, therefore, consistent with the trend of performance *previous* to experimentation.

(2) The accomplishment quotients of the two groups would be expected to differ if the teachers differ in influence upon achieve-

ment. Table II presents the A.Q.'s of the comparative groups at four different periods during experimentation. The fact is that there is no reliable difference between the groups as regards A.Q. Such apparent difference as there is favors the teacher of the Lower Group.

TABLE II.—ACCOMPLISHMENT QUOTIENTS OF THE TWO GROUPS DURING EXPERIMENTATION, ACCORDING TO THE STANFORD ACHIEVEMENT TEST

Date	A. Q. Higher Group	A. Q. Lower Group
Sept., 1923	94.5	91.0
June, 1924	96.5	99.0
Sept., 1924	98.0	100.0
June, 1925	96.5	99.5
Average	96.4	97.4

(3) Furthermore, in September, 1924, nine of the forty pupils here involved were shifted from one classroom teacher to the other (a few very young children of higher I.Q. were moved to the teacher of the Lower Group, and a few older children of lower I.Q. were moved to the teacher of the Higher Group). Analysis of the performance of these children after shifting shows no consistent departure from the trend for the years preceding. These facts are presented in Table III in the form of A.Q.'s for these nine individuals for a period of two years.

TABLE III.—THE A. Q.'S OF INDIVIDUALS SHIFTED FROM ONE TEACHER TO ANOTHER IN SEPTEMBER, 1924

(The first three cases in the table were moved from the teacher of the Higher Group to the teacher of the Lower Group. The other six were moved from the Lower to the Higher. The first two cases in the table belong to our Higher Group, the other seven to our Lower Group.)

Child	I.Q.	A. Q. (Stanford Achievement Test)			
		Sept., 1923	June, 1924	Sept., 1924	June, 1925
A	159	100	104	105	104
B	152	95	97	102	101
C	156	101	100	104	104
D	149	89	93	90	96
E	145	102	101	106	106
F	141	100	104	106	101
G	142	97	100	100	100
H	139	89	97	99	99
I	145	97	99	103	99

(4) It would be difficult to show why, if the influence of the two teachers in question were the cause of differences in achievement, that influence should affect some mental functions, but not others. For instance, it would be hard to explain why it should show itself sharply in reading, but not in multiplication of whole numbers; in spelling, but not in addition of whole numbers.

From this evidence we infer that the influence of the two classroom teachers did not interfere in any appreciable way with the equalization of educational opportunity. While it is no doubt true that for *individual* pupils educational opportunity from birth may have been somewhat unequal, for the two *matched groups* it was surely equal in all essential and important respects.

Home Conditions

Home conditions were rated by Miss Cobb. Rating was according to the Whittier Scale, with 25 points as a maximal score for a home. Table IV, showing the comparative home ratings for our 20 pairs, reveals the excellence of the homes from which most of these very gifted children come.[3]

Two of the children in our Higher Group have poorer homes than any child in the Lower Group. These two homes are decidedly 'unprivileged,' located in slum districts, crowded, and devoid of books.

Policy of Instruction

The policy of instruction was to let each child learn at his own natural rate. Adjustments were made constantly according to individual advancement in any subject of study or project covering the prescribed curriculum, and also including in the daily pro-

---

[3] As the Whittier Scale does not give scope for a very wide range of home-ratings, it does not seem surprising that it should fail to discriminate among homes of children all of whom fall in the top one half of one percent as regards I.Q. Our data show only that our 146 I.Q. group has as many homes *at or near the maximum* for excellence as does the 165 I.Q group, by the criteria of the Whittier Scale.

We were not surprised, in fact, that the ratings (made independently of the final composition of our two groups) should have yielded no discrimination among the homes of children so homogeneous as to parentage that *all* rate in one half of a given percentile as concerns I.Q. We think it very probable that children limited to one half of *any* percentile on the distribution of I.Q would show homes rating alike.

TABLE IV.—COMPARATIVE HOME RATINGS FOR 20 EXPERIMENTAL PAIRS, OF HIGHER AND LOWER GROUPS, RESPECTIVELY, ACCORDING TO THE WHITTIER SCALE FOR RATING HOMES

Pair	Home Rating H	Home Rating L	Pair	Home Rating H	Home Rating L
1.	25	25	11.	24	24
2.	25	25	12.	17	24
3.	25	23	13.	25	25
4.	25	25	14.	25	23
5.	24	24	15.	23	22
6.	25	24	16.	17	20
7.	25	24	17.	24	25
8.	23	23	18.	23	24
9.	20	20	19.	21	25
10.	25	25	20.	25	21
			Mean	23.30	23.25

gram much unusual opportunity for education. French, social science, the history of civilization, music, biography, physical training, were all introduced to enrich the curriculum. These opportunities were open to all alike.

Under these conditions, the children learned as much as they would or could, and were tested at intervals to ascertain their achievement. Whenever individual difficulties were detected, help was given. All had equal access to this assistance.

TESTS OF ACHIEVEMENT

Standardized tests of achievement were administered by Miss Cobb from time to time over the period of three years. Different equivalent forms of each test, where such exist, were employed on consecutive occasions. All children were tested by the same form on any given occasion. Every test was given and scored in exactly the prescribed manner, so that no further description of the testing is necessary.

In such prolonged experimentation it is inevitable that a few children will be absent on some occasions of testing. It was impossible to have all of these absentees make up every test immediately. Thus, a few records had to be interpolated. This was done, after the first three trials, by giving each absentee the highest score which he had reached (or bettered) three times previously.

For example, if a child, upon being absent, had three best scores of 6, 5, 6, on previous occasions, he was given a score of 5 on the occasion of his absence; one having best previous scores of 4, 4, 7, would be given a score of 4.

Between 5 and 10 percent of scores in *each* group were thus inferred. As any error which might arise from this procedure would affect both groups equally, it could not exercise an invalidating influence upon the results.

On the Stanford Achievement Tests, however, there are no interpolated scores whatever. If a child was absent on the occasion of testing, he was followed up immediately and tested, with the exception of one child, in June, 1924. On this single occasion, the median of 19 instead of 20 children was found in the Lower Group (to which this absent child belonged).

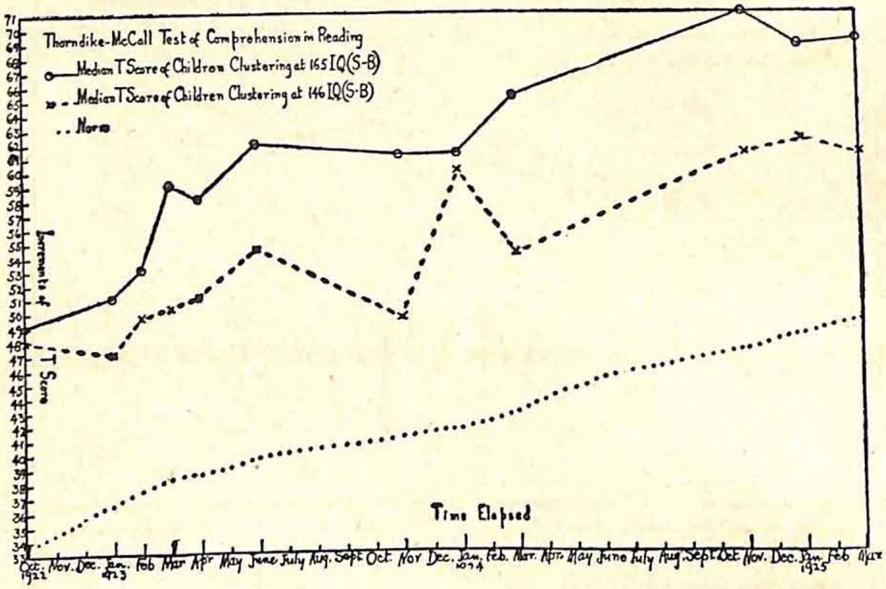
Results

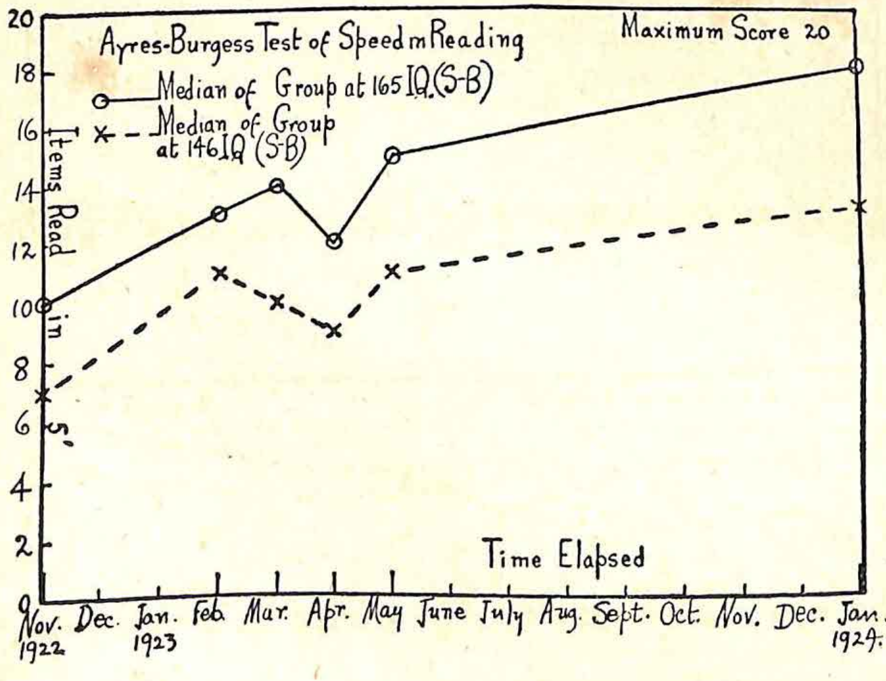
The results of experimentation are presented by means of tables and graphs which seem to require no verbal description. In a few of the tests given, progress is not measured on the same scale, because different kinds of tests were used from occasion to occasion. In these cases the results are presented in Table V. All other tests represent progressive achievement, each on the same scale as a base from time to time, and are presented in the form of graphs. It will be noted that results from the miscellaneous achievements presented in Table V are in agreement with the data presented in the graphs, in that they show that the group testing at 165 I.Q. excels the group testing at 146 I.Q. in intellectual performances.

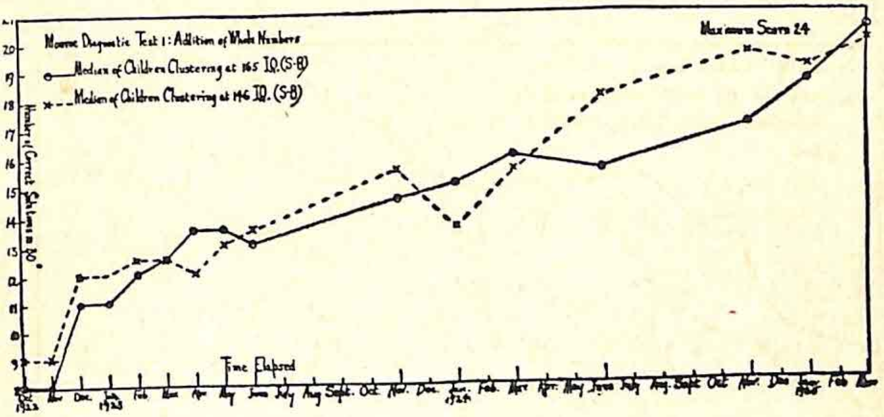
Attention should be directed especially to the results from the tests of *penmanship.* There is no reliable difference between our groups in this achievement. It is safe to say, also, that they do not excel unselected children of their age in speed of penmanship, and excel them but slightly, if at all, in quality.

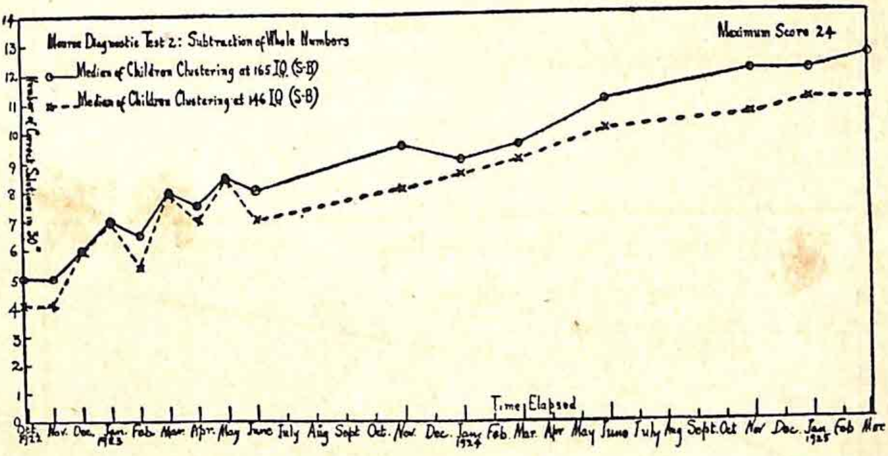
TABLE V.—COMPARISON OF HIGHER AND LOWER GROUPS WHERE DATA WERE INSUFFICIENT FOR GRAPHS

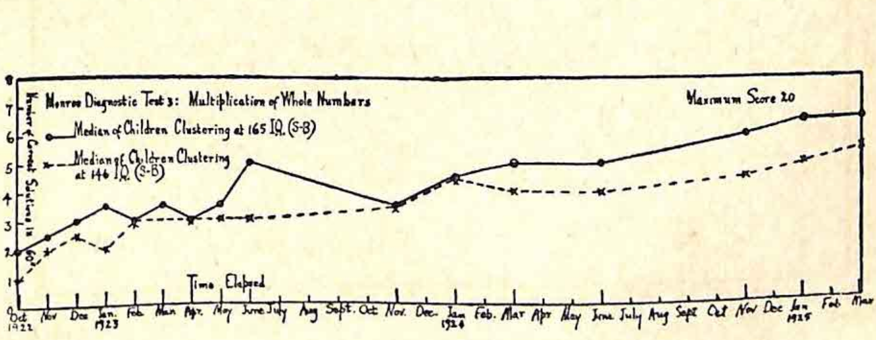
Test Given	Group	1922			1923							1924							1925		
		Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	June	Nov.	Jan.	Feb.	Mar.	Apr.	June	Nov.	Dec.	Jan.	Mar.	Apr.
Chapman-Cook Speed of Reading	Higher																		22.5		
	Lower																		20		
I. E. R. Graded Word Knowledge	Higher											70.5									
	Lower											59									
Stanford-Binet Vocabulary (Approx. time of test varied)	Higher	41						46							54.25						61
	Lower	34.5						40.75							49.5						
Spelling (Buckingham-Ayres) (Difficulty varied)	Higher	88	78	92	92	88	76	84	82	92	96	80		86		92	79		65		
	Lower	78	72	80	70	80	54	78	76	69	80	70		68		88	54		37		
Spelling (Horn-Ashbaugh)	Higher																			92	
	Lower																			80	
Composition (Nassau Scale)	Higher	4.15		4.9	5.6																
	Lower	3.4		3.6	4.1																
Composition (Willing Scale)	Higher								62.5		70	65		75							
	Lower								51		55	55		64							
Penmanship (N. Y. C. Scale) Quality	Higher					50	50	50	42		50	52									
	Lower					50	55	48	58.5		46.5	49.5									
Penmanship Rate (in terms of grade level)	Higher		4		4a	5a	5	4b	4a		5b	5a									
	Lower		3		4a	5b	5b	5b	4b-5a		5a	7a									
Van Wagenen History Information	Higher												47			69.9		errors 12.25			
	Lower												40			70.7		14.5			
Van Wagenen History Information	Higher												46.5			82.75		18			
	Lower												39.5			78		20.5			
Van Wagenen History Judgment	Higher												53.5								
	Lower												41.5								
Van Wagenen History Character	Higher												36								
	Lower												28								
Monroe Reasoning Test in Arithmetic (Total Parts I, II, and III)	Higher								31.5			39.5									
	Lower								23			26									

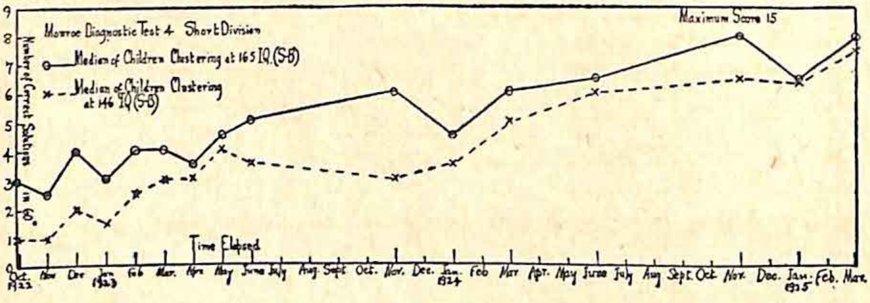

Thorndike-McCall Test of Comprehension in Reading
Median T Score of Children Clustering at 165 IQ (S-B)
Median T Score of Children Clustering at 146 IQ (S-B)
Norm
Increments of T Score
Time Elapsed
Oct. 1922
Nov.
Dec.
Jan. 1923
Feb
Mar
Apr
May
June
July
Aug.
Sept.
Oct.
Nov
Dec.
Jan. 1924
Feb.
Mar.
Apr.
May
June
July
Aug.
Sept.
Oct.
Nov.
Dec.
Jan. 1925
Feb

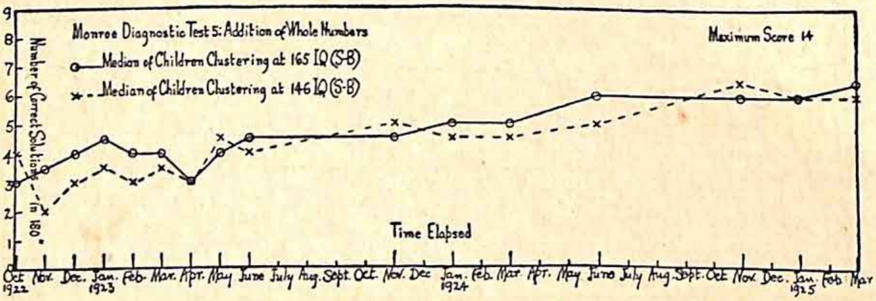

Ayres-Burgess Test of Speed in Reading
Maximum Score 20
Median of Group at 165 IQ (S-B)
Median of Group at 146 IQ (S-B)
Items Read in 5'
Time Elapsed
20
18
16
14
12
10
8
6
4
2
0
Nov. 1922
Dec.
Jan. 1923
Feb.
Mar.
Apr.
May
June
July
Aug.
Sept.
Oct.
Nov.
Dec.
Jan. 1924.

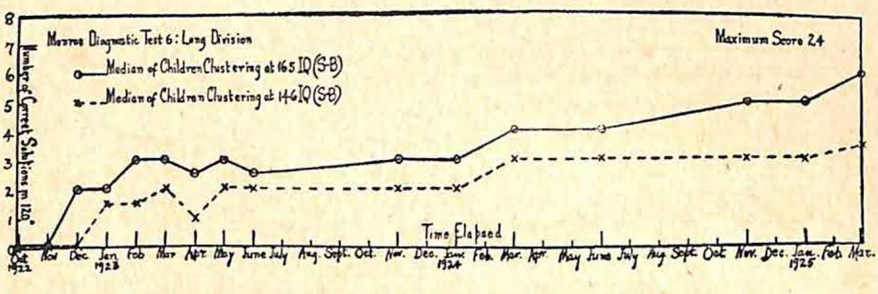

Monroe Diagnostic Test 1: Addition of Whole Numbers
Median of Children Clustering at 165 I.Q. (S-B)
Median of Children Clustering at 146 I.Q. (S-B)
Maximum Score 24
Time Elapsed

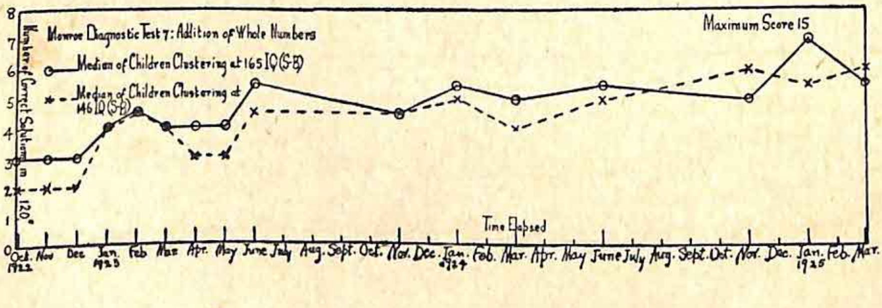

Monroe Diagnostic Test 2: Subtraction of Whole Numbers
Median of Children Clustering at 165 I.Q. (S-B)
Median of Children Clustering at 146 I.Q. (S-B)
Maximum Score 24
Time Elapsed

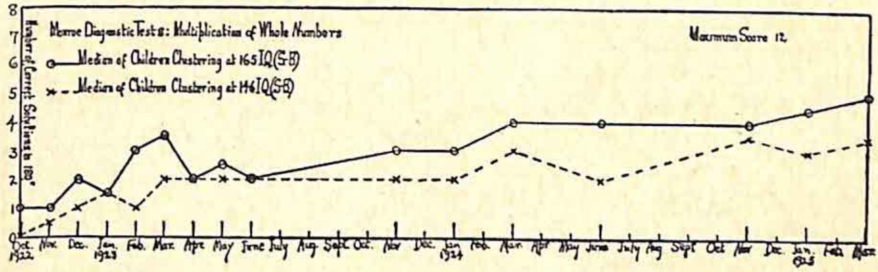

Monroe Diagnostic Test 3: Multiplication of Whole Numbers
Median of Children Clustering at 165 I.Q. (S-B)
Median of Children Clustering at 146 I.Q. (S-B)
Maximum Score 20
Time Elapsed

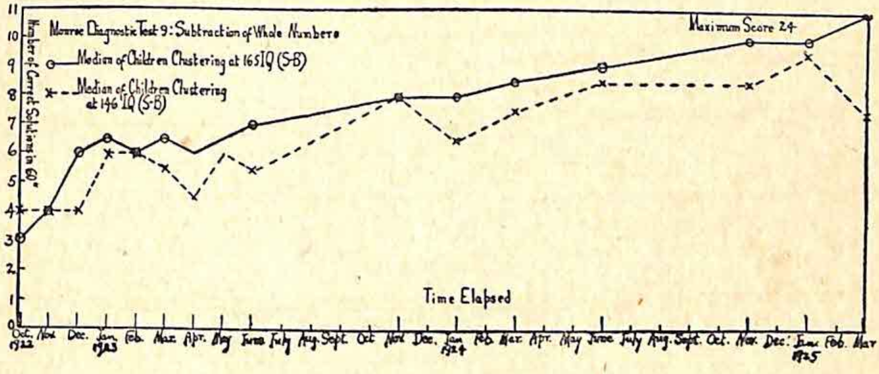

Monroe Diagnostic Test 4 Short Division
Maximum Score 15
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 180"
Time Elapsed

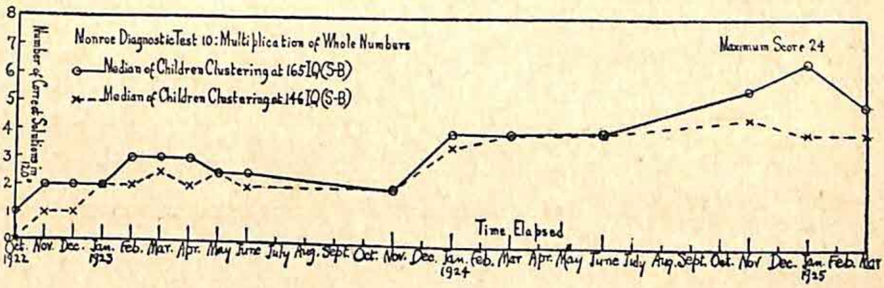

Monroe Diagnostic Test 5: Addition of Whole Numbers
Maximum Score 14
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 180"
Time Elapsed

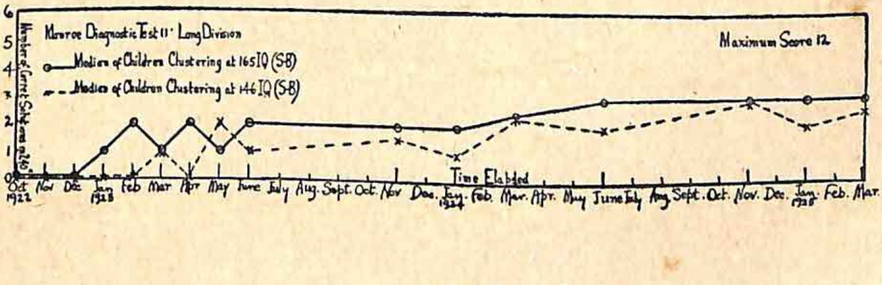

Monroe Diagnostic Test 6: Long Division
Maximum Score 24
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 120"
Time Elapsed

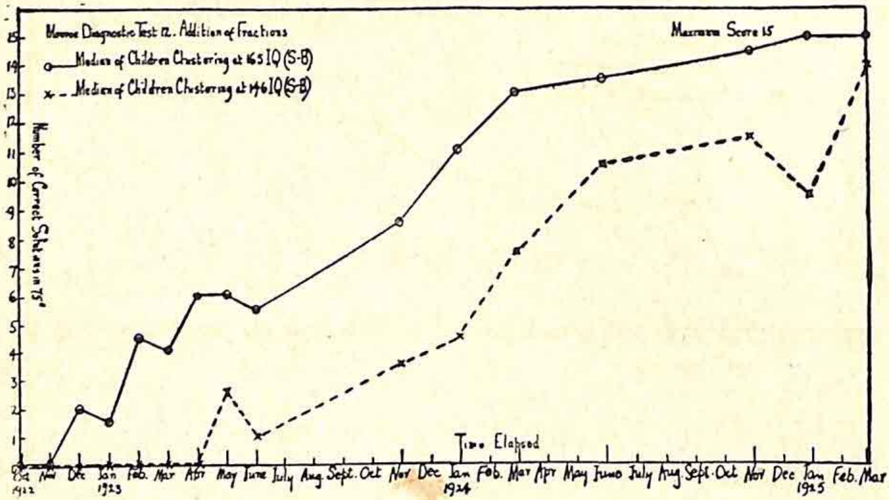

Monroe Diagnostic Test 7: Addition of Whole Numbers
Maximum Score 15
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 120"
Time Elapsed

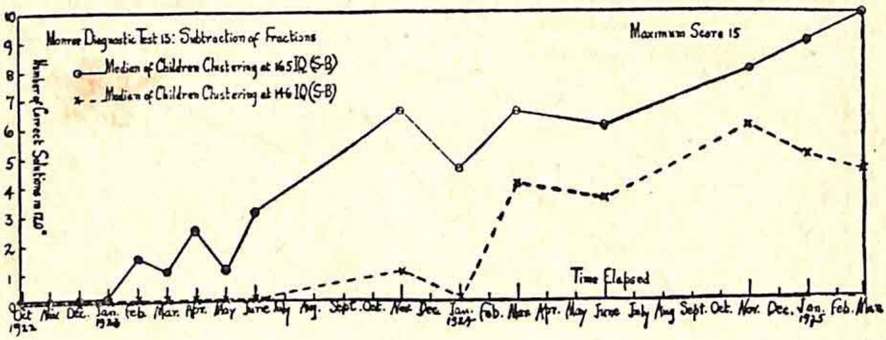

Monroe Diagnostic Test 8: Multiplication of Whole Numbers
Maximum Score 12
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 180"

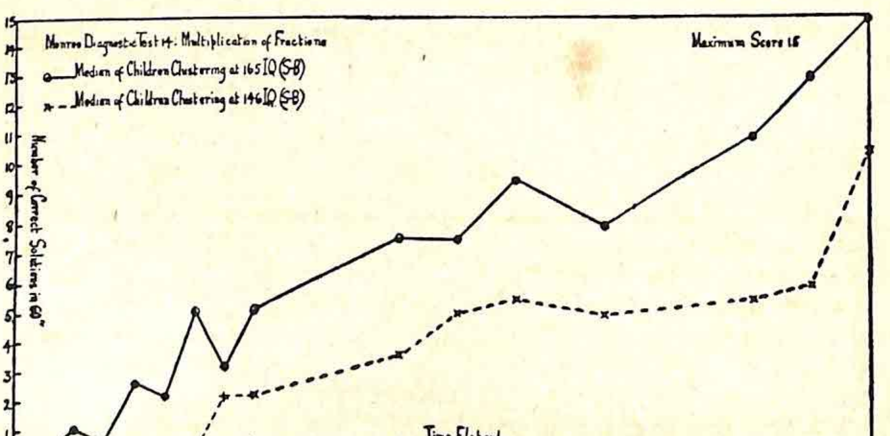

Monroe Diagnostic Test 9: Subtraction of Whole Numbers
Maximum Score 24
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 60"
Time Elapsed

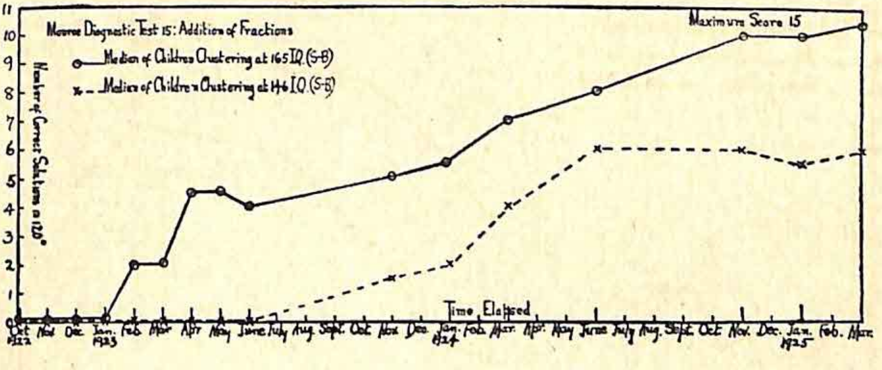

Monroe Diagnostic Test 10: Multiplication of Whole Numbers
Maximum Score 24
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 120"
Time Elapsed

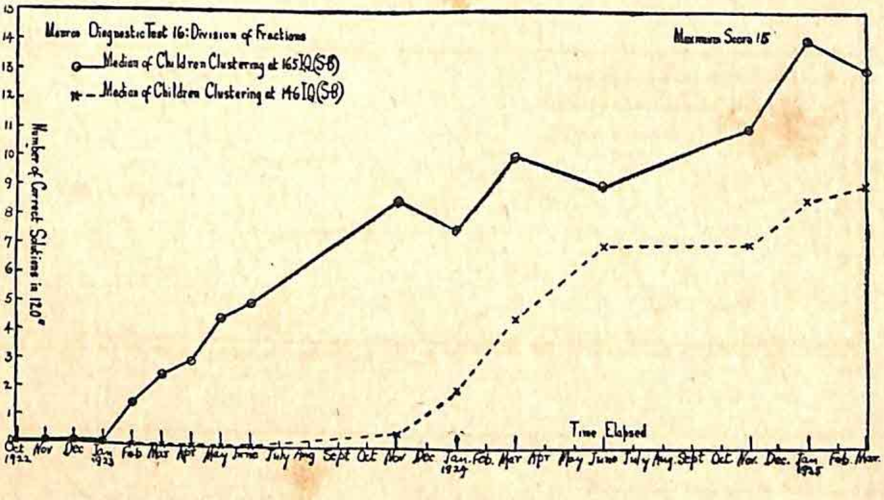

Monroe Diagnostic Test 11: Long Division
Maximum Score 12
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Time Elapsed

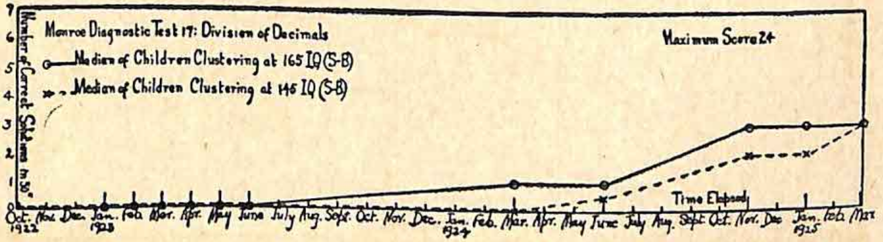

Monroe Diagnostic Test 12. Addition of Fractions
Maximum Score 15
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 75"
Time Elapsed
Oct Nov Dec Jan Feb. Mar Apr May June July Aug. Sept. Oct Nov Dec Jan Feb. Mar Apr May June July Aug. Sept. Oct Nov Dec Jan Feb. Mar
1922 1923 1924 1925

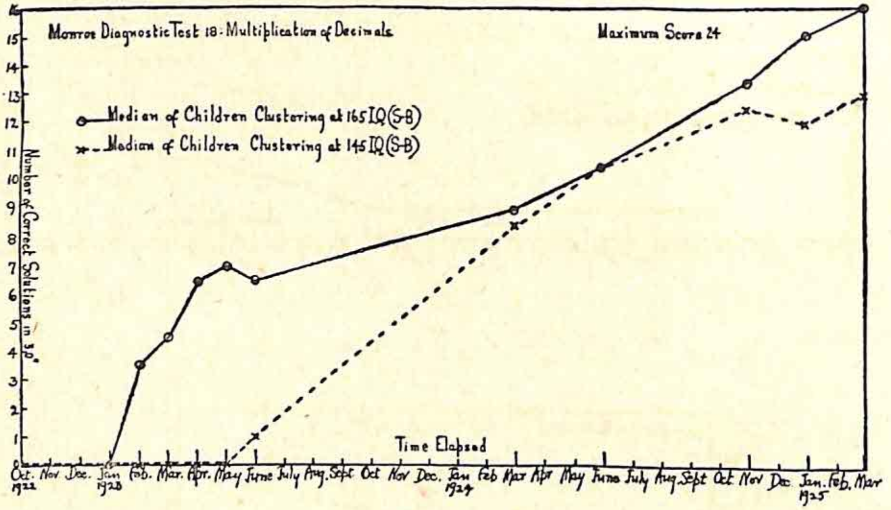

Monroe Diagnostic Test 13: Subtraction of Fractions
Maximum Score 15
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 120"
Time Elapsed
Oct Nov Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov Dec. Jan. Feb. Mar Apr. May June July Aug Sept. Oct. Nov. Dec. Jan. Feb. Mar
1922 1923 1924 1925

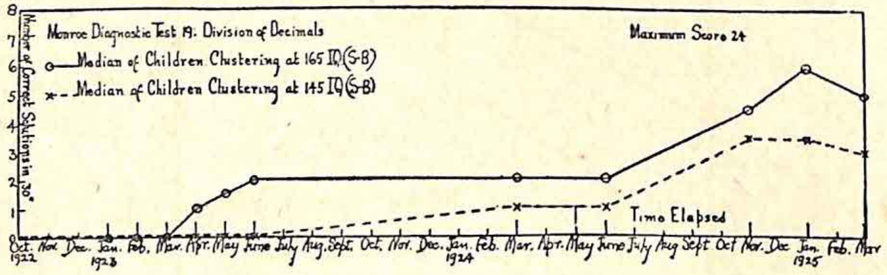

Monroe Diagnostic Test 14: Multiplication of Fractions
Maximum Score 15
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 146 IQ (S-B)
Number of Correct Solutions in 60"
Time Elapsed
Oct. Nov Dec Jan. Feb. Mar. Apr. May June July Aug Sept Oct. Nov Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept Oct Nov Dec Jan. Feb. Mar
1922 1923 1924 1925

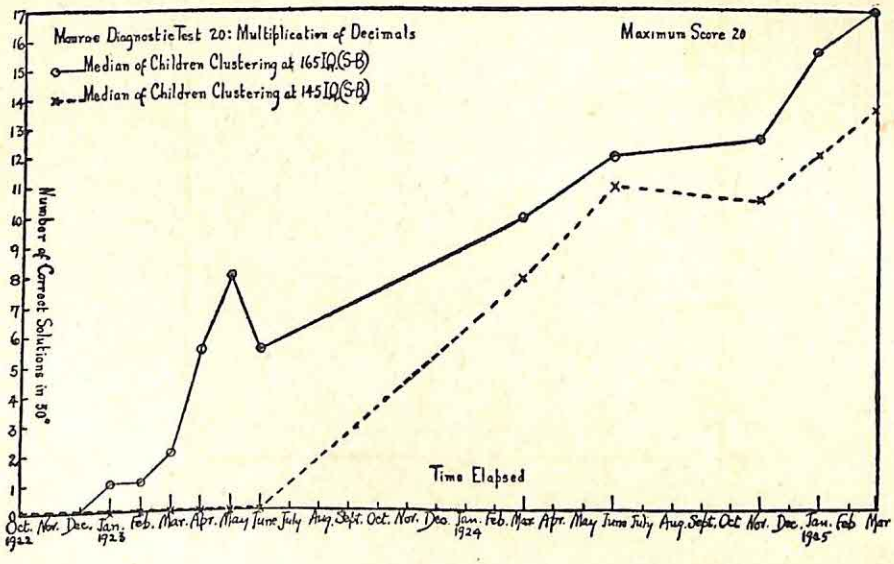

Monroe Diagnostic Test 15: Addition of Fractions
Maximum Score 15
Median of Children Clustering at 165 I.Q. (S-B)
Median of Children Clustering at 146 I.Q. (S-B)
Number of Correct Solutions in 120"
Time Elapsed
Oct. 1922 Nov. Dec. Jan. 1923 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. 1924 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. 1925 Feb. Mar.

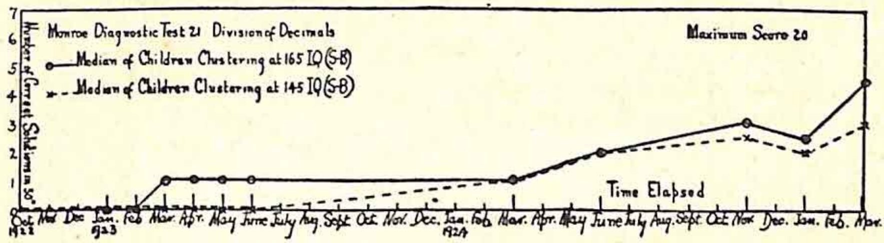

Monroe Diagnostic Test 16: Division of Fractions
Maximum Score 15
Median of Children Clustering at 165 I.Q. (S-B)
Median of Children Clustering at 146 I.Q. (S-B)
Number of Correct Solutions in 120"
Time Elapsed
Oct. 1922 Nov. Dec. Jan. 1923 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. 1924 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. 1925 Feb. Mar.

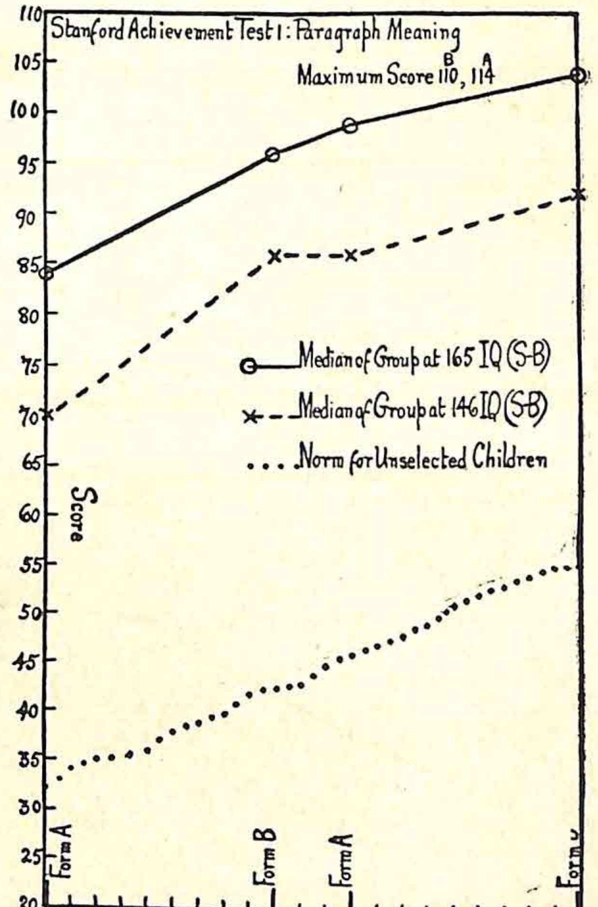

Monroe Diagnostic Test 17: Division of Decimals
Maximum Score 24
Median of Children Clustering at 165 I.Q. (S-B)
Median of Children Clustering at 146 I.Q. (S-B)
Number of Correct Solutions in 30"
Time Elapsed
Oct. 1922 Nov. Dec. Jan. 1923 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. 1924 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. 1925 Feb. Mar.

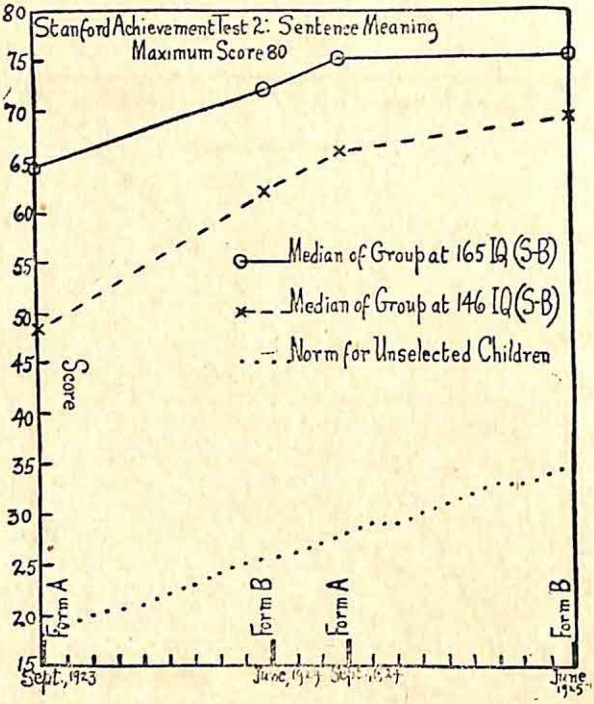

Monroe Diagnostic Test 18: Multiplication of Decimals
Maximum Score 24
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 145 IQ (S-B)
Number of Correct Solutions in 30"
Time Elapsed
Oct. Nov. Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. Feb. Mar.
1922 1923 1924 1925

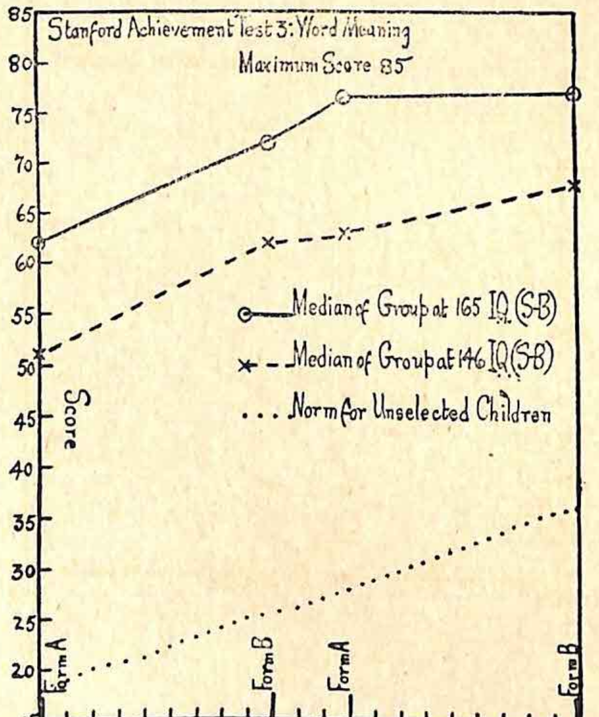

Monroe Diagnostic Test 19: Division of Decimals
Maximum Score 24
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 145 IQ (S-B)
Number of Correct Solutions in 30"
Time Elapsed
Oct. Nov. Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. Feb. Mar.
1922 1923 1924 1925

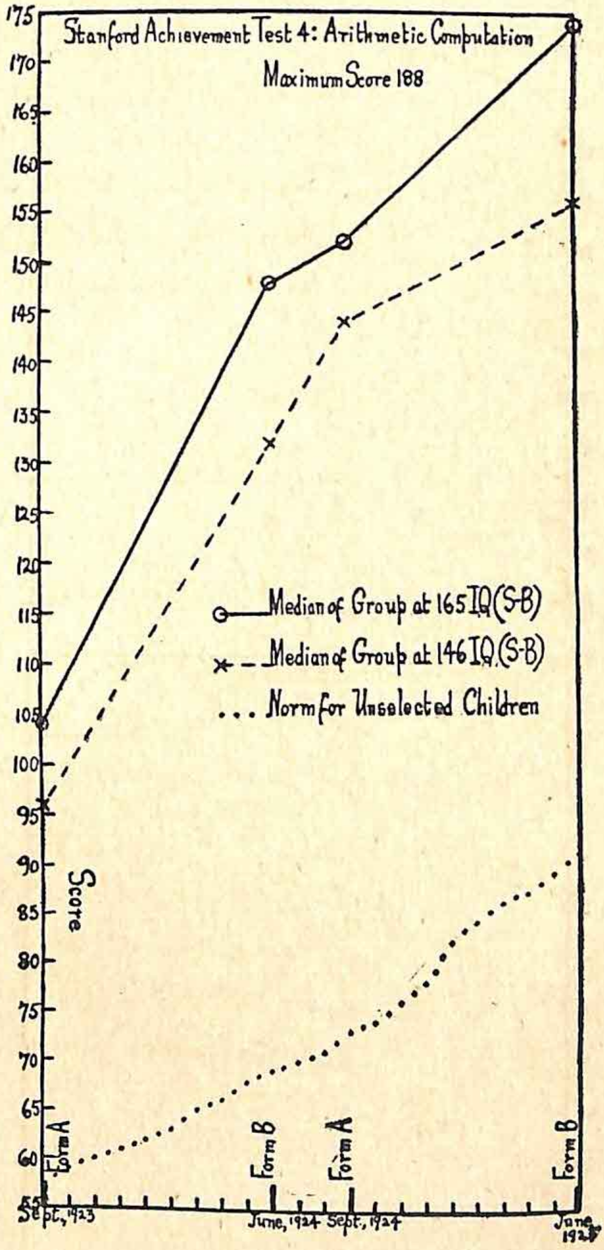

Monroe Diagnostic Test 20: Multiplication of Decimals
Maximum Score 20
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 145 IQ (S-B)
Number of Correct Solutions in 50"
Time Elapsed
Oct. Nov. Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. Feb. Mar.
1922 1923 1924 1925

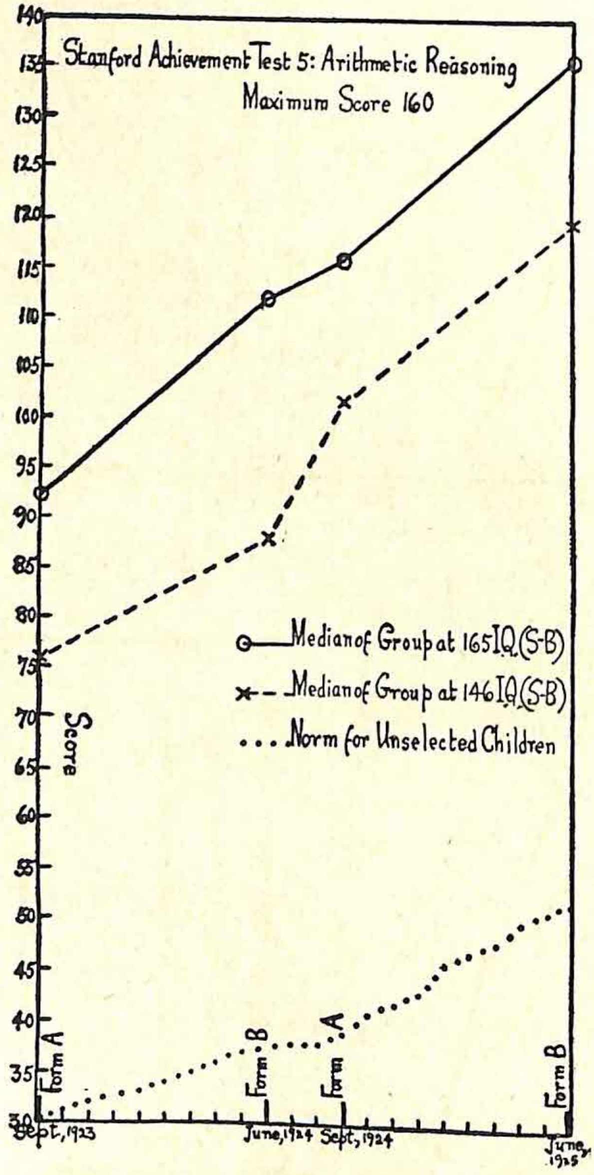

Monroe Diagnostic Test 21 Division of Decimals
Maximum Score 20
Median of Children Clustering at 165 IQ (S-B)
Median of Children Clustering at 145 IQ (S-B)
Time Elapsed
Oct. 1922 Nov. Dec. Jan. 1923 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. 1924 Feb. Mar. Apr. May June July Aug. Sept. Oct. Nov. Dec. Jan. Feb. Mar.
0 1 2 3 4 5 6 7

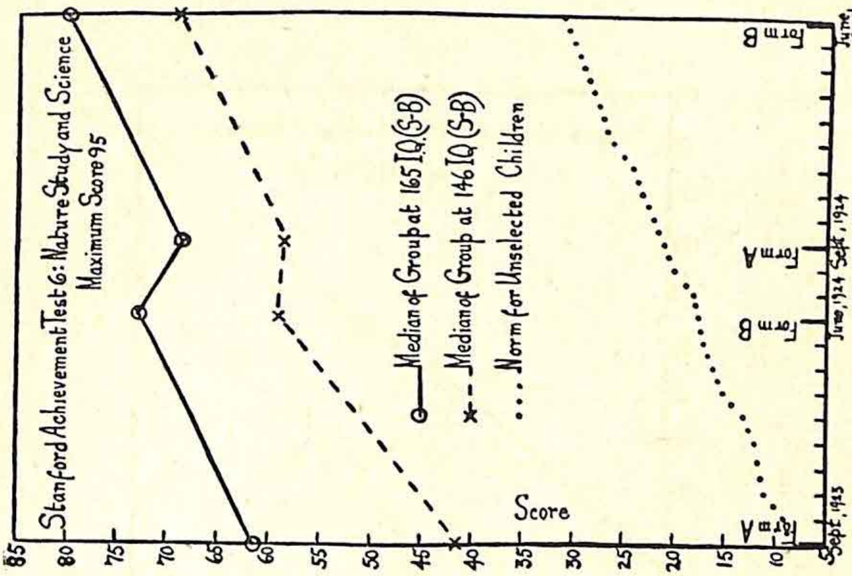

Stanford Achievement Test 1: Paragraph Meaning
Maximum Score 110 B, 114 A
Median of Group at 165 IQ (S-B)
Median of Group at 146 IQ (S-B)
Norm for Unselected Children
Score
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110
Form A
Form B
Form A
Form B
Sept., 1923
June, 1924 Sept., 1924
June, 1925

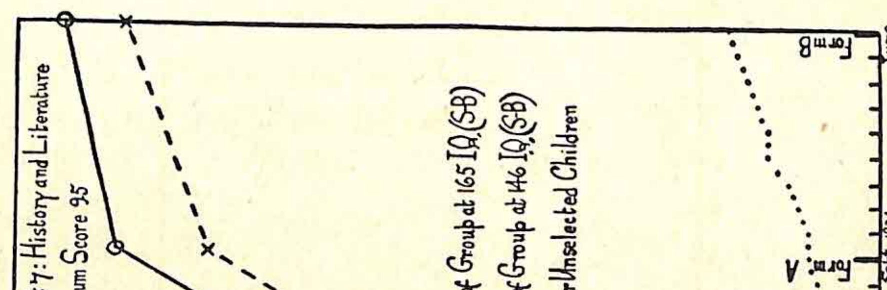

Stanford Achievement Test 2: Sentence Meaning
Maximum Score 80
Median of Group at 165 IQ (S-B)
Median of Group at 146 IQ (S-B)
Norm for Unselected Children
Score
Form A
Form B
Form A
Form B
Sept., 1923
June, 1924
June 1925
80
75
70
65
60
55
50
45
40
35
30
25
20
15

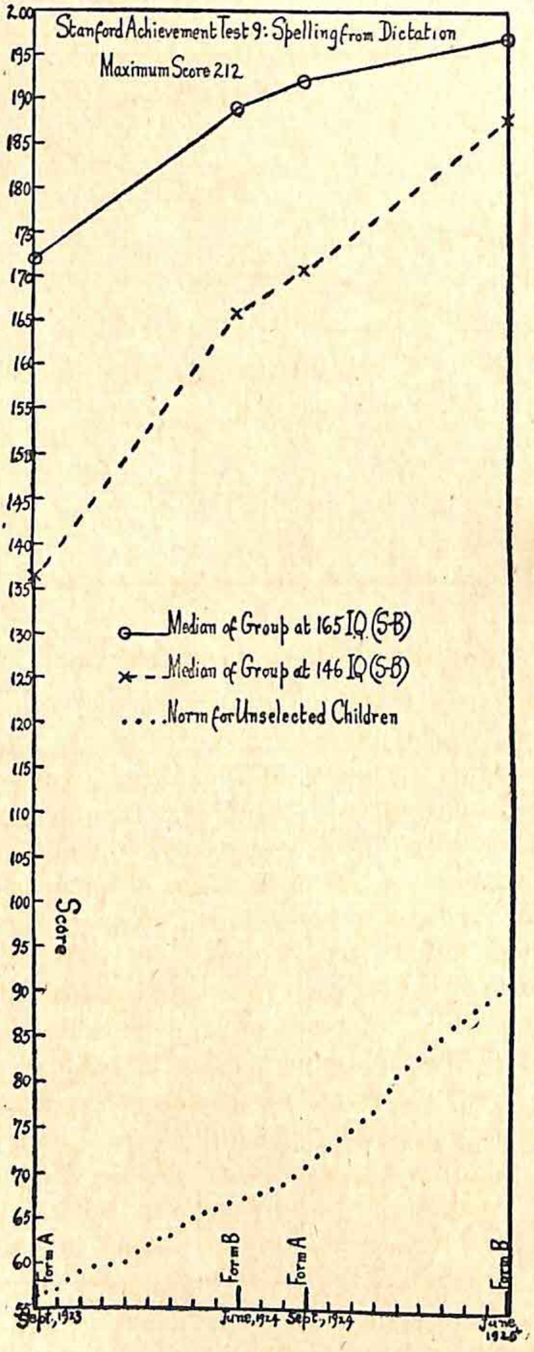

Stanford Achievement Test 3: Word Meaning
Maximum Score 85
Median of Group at 165 IQ (S-B)
Median of Group at 146 IQ (S-B)
Norm for Unselected Children
Score
Form A
Form B
Form A
Form B
Sept., 1923
June, 1924
Sept., 1924
June, 1925
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15

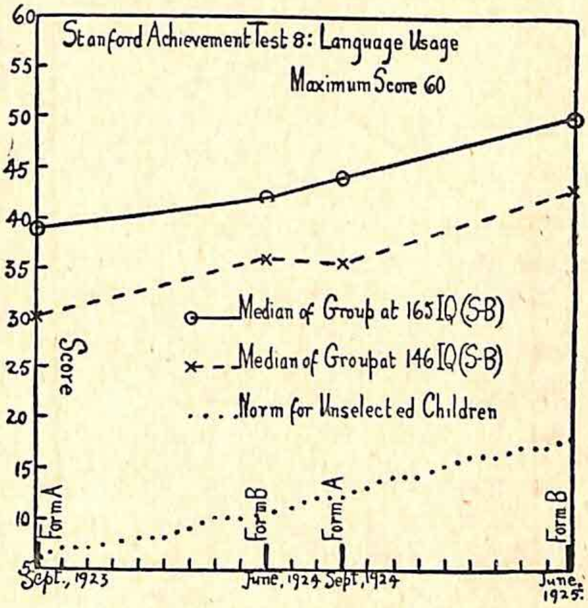

Stanford Achievement Test 4: Arithmetic Computation
Maximum Score 188
Median of Group at 165 IQ (S-B)
Median of Group at 146 IQ (S-B)
Norm for Unselected Children
Score
175
170
165
160
155
150
145
140
135
130
125
120
115
110
105
100
95
90
85
80
75
70
65
60
55
Form A
Form B
Form A
Form B
Sept., 1923
June, 1924
Sept., 1924
June 1925


Stanford Achievement Test 5: Arithmetic Reasoning
Maximum Score 160
Median of Group at 165 I.Q. (S-B)
Median of Group at 146 I.Q. (S-B)
Norm for Unselected Children
Score
140
135
130
125
120
115
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
Form A
Form B
Form A
Form B
Sept., 1923
June, 1924
Sept., 1924
June, 1925


Stanford Achievement Test 6: Nature Study and Science
Maximum Score 95
Median of Group at 165 I.Q. (S-B)
Median of Group at 146 I.Q. (S-B)
Norm for Unselected Children
Score
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
Form A
Form B
Form A
Form B
Sept. 1923
June, 1924
Sept. 1924
June, 1925


Stanford Achievement Test 7: History and Literature
Maximum Score 95
Median of Group at 165 I.Q. (S-B)
Median of Group at 146 I.Q. (S-B)
Norm for Unselected Children
Score
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
Form A
Form B
Form A
Form B
Sept. 1923
June, 1924
Sept. 1924
June, 1925


Stanford Achievement Test 9: Spelling from Dictation
Maximum Score 212
Median of Group at 165 I.Q. (S-B)
Median of Group at 146 I.Q. (S-B)
Norm for Unselected Children
Score
200
195
190
185
180
175
170
165
160
155
150
145
140
135
130
125
120
115
110
105
100
95
90
85
80
75
70
65
60
55
Form A
Form B
Form A
Form B
Sept., 1923
June, 1924
Sept., 1924
June, 1925



Interpretation of the Results

It is immediately evident that, among the functions herein tested, comparative achievement varies greatly. In a few functions there is no advantage from the additional intelligence represented by 165 I.Q., as compared with 146 I.Q. In other functions the difference in favor of the Higher Group seems comparatively slight, but in most instances the difference seems pronounced.

The question arises as to how we may compare the magnitudes of the differences found; how we may, for instance, compare a difference of one correct solution in Long Division (Monroe Diagnostic Test 11) with a difference of ten points in T score in the Thorndike-McCall Test of comprehension in reading.

We have adopted the following method of arriving at a rough estimate of the magnitudes of the differences found between the two compared groups. The *time saved* by the Higher Group in arriving at the point reached as a *final score* by the Lower Group has been calculated in months, and the *saving* thus accomplished has been taken as a sign of *the magnitude of the difference.*

These calculations are but approximate, for several reasons. In the first place, it must be borne in mind that the testing of performance was discontinuous, so that we cannot tell just what the respective

achievements of both groups were at every moment in time. We know these only for certain sample moments, as indicated on the graphs. Therefore, to find the point in time at which one group equals the final achievement of the other, we have to calculate from the *slant* of the curve between two tested moments. This, of course, is not so accurate as the result from continuous testing would be. In the second place, there may arise in a few of the tests an error of handicap upon the Higher Group, from the fact that children having the highest I.Q.'s reached a top or nearly top score in some of the tests before the final trials. This error we have sought to avoid by our method of comparison which cites the Higher Group at *any* point in its career wherever it meets the final median score of the Lower Group. We have sought also to avoid it by taking the medians rather than the means of the groups as points of comparison.

Similar comparisons of 'time saved' with the norms for unselected children, where such are available, cannot be made, because neither of our experimental groups overlaps at all with unselected children during the time covered by the three years of experimentation. In the Stanford Achievement Tests, for instance, our two groups at 9 years, 6 months, of age very much exceed the scores of ordinary children who are 11 years, 3 months, old (the age at which our experimental groups were tested for the last time). We have, however, included on the graphs the normal curves for unselected children in achievements for which age norms are available or can be inferred.

Table VI shows how the various achievements compare with each other in the differences which they show between children clustering at 165 I.Q. and those clustering at 146 I.Q. according to the method of calculating *time saved,* which we have just described. The tests are listed in order of amount of time saved by the Higher Group, with those listed first in which most time was saved. It is interesting that the Higher Group saved time in every test of achievement but one, though the amounts vary from 16 months to 0 months in various performances.

In those functions or tasks which head the list for magnitude of difference between our groups, the curves of performance are never in contact from the time either group rises above zero. In those which fall at the bottom of the list there is practically no advantage realized by the Higher Group from its additional + 2 P.E. of intellect. Addition of Whole Numbers can be performed as well by

TABLE VI.—TIME SAVED BY THE HIGHER I. Q. GROUPS IN ATTAINING VARIOUS LEVELS OF ACHIEVEMENTS

Test of Achievement	Months Saved by Higher Group
Word Meaning, Stanford Achievement 3	16.0
Paragraph Meaning, Stan. Ach. 1	15.0
Sentence Meaning, Stan. Ach. 2	14.5
Nature Study and Science, Stan. Ach. 6	14.5
*Understanding of Paragraphs, Thorndike-McCall	13.5
*Addition of Fractions, Monroe Diagnostic Test 9	13.5
Long Division, Mon. Diag. 6	13.0
Multiplication of Whole Numbers, Mon. Diag. 8	13.0
Spelling from Dictation, Stan. Ach. 9	12.5
Language Usage, Stan. Ach. 8	11.0
History and Literature, Stan. Ac. 7	9.0
*Long Division, Mon. Diag. 11	9.0
Subtraction of Whole Numbers, Mon. Diag. 2	9.0
Division of Fractions, Mon. Diag. 16	9.0
*Subtraction of Fractions, Mon. Diag. 13	9.0
Speed in Silent Reading, Ayres-Burgess	8.5
Arithmetic Computation, Stan. Ach. 4	7.5
*Multiplication of Whole Numbers, Mon. Diag. 10	7.5
Arithmetic Reasoning, Stan. Ach. 5	7.0
Multiplication of Whole Numbers, Mon. Diag. 3	7.0
Addition of Fractions, Mon. Diag. 12	6.5
*Subtraction of Whole Numbers, Mon. Diag. 9	6.5
*Division of Decimals, Mon. Diag. 19	6.0
Short Division, Mon. Diag. 4	6.0
Multiplication of Decimals, Mon. Diag. 18	5.0
Multiplication of Fractions, Mon. Diag. 14	5.0
Division of Decimals, Mon. Diag. 17	4.0
Division of Decimals, Mon. Diag. 21	4.0
Multiplication of Decimals, Mon. Diag. 20	3.5
Addition of Whole Numbers, Mon. Diag. 7	3.0
Addition of Whole Numbers, Mon. Diag. 1	0.5
*Addition of Whole Numbers, Mon. Diag. 5	0.0

*In the instances where the final score of the Lower Group is not its best score, the comparison has been made on the basis of the best score as a final score.

children of 146 I.Q. as by children of 165 I.Q., at the ages studied. How is this to be explained?

The failure of the Higher Group to draw away from the Lower Group in these achievements cannot be explained by saying that both groups had reached the maximum of proficiency, for at no point until near the end is there approach to a perfect score. The curve rises quite steadily from one occasion of testing to the next, but rises equally for both groups.

Perhaps the explanation of these facts lies in the motor mechanics of the operation. It probably is true that young children of I.Q. 146 can perform additions with whole numbers, like those

found in Monroe Tests 7, 5 and 1, about as rapidly as the eye and hand can execute the task; and that at and above this level, therefore, increments of intellect give no advantage. Both groups of very intelligent children show, however, a decided advantage over unselected children, even in these operations; for Monroe's norms are well below the performances here displayed, according to grade norms (and inferred age norms), both when children enter school at five years and when they enter at six years.[3] The point where such achievement ceases to differentiate degrees of intellect falls, therefore, somewhere between 100 I.Q. and 146 I.Q.

What is meant by the interpretation here offered may be further illustrated by a very simple example. Ten-year-olds of I.Q. 100 can tell their names as readily as can those of I.Q. 150. So thoroughly proficient has each group become in this very simple task that each is limited in performance of it only by the mechanics of articulation, which do not differentiate intellectual groups of ten-year-olds above a certain very low level.

In appraising all differences it is necessary of course to bear in mind the *maximal possible score*. This is given on the graphs. It will be noted that in some performances the Lower Group appears to approach the Higher Group toward the end of the comparison; whereas inspection of the data will show that so many of the Higher Group have in fact reached or so nearly reached the highest possible proficiency that the comparison is no longer wholly valid, because the Higher Group has exhausted opportunity. Herein lies one of the most important criticisms of available tests for purposes of experimentation such as this. Children as gifted as the best in our Higher Group, that is of I.Q. above 160, are 'off scale' by the time they are ten or eleven years old. We have tried to allow as well as we could for this handicap upon our Higher Group by adopting group medians, instead of group means, as our basis of comparison.

Speculative Interpretation

From our series of achievements as they stand we infer that a difference of +2 P.E. in I.Q. above 146 I.Q. (S-B) at the ages studied shows itself in comparative performance in proportion to the complexity and difficulty of the task attempted.

[3] We have been unable to gain precise information as to age norms for the Monroe Diagnostic Tests.

The few performances in which there is no demonstrable difference between our two groups are relatively simple, as is shown by the fact that they have fewer elements and thus require fewer associations than tests which differentiate the groups. For instance, in addition of fractions (which clearly shows an advantage for the group at 165 I.Q.) there are all of the elements involved in addition of whole numbers (which gives them no advantage) and *many more besides.* The former requires all the associations required by the latter, and other groups of associations as well.

Also, it is true that average children can make a score above zero, at all the ages here studied, in addition of *whole numbers,* whereas in addition of *fractions* they do not rise above zero until they are about eleven years old. It may pertinently be said that average children are not taught fractions until that age, but is there not some reason for this?

We infer that it would be possible to discover by experimentation scores of tasks which, because of their comparative simplicity, would not differentiate 'very high' from 'still higher' intelligence. Experimentation with groups like ours affords a method of discovering which among the world's tasks will clearly give advantage to the best one tenth of one percent of intellects. Among tasks tried by us the following seem to give most advantage to increments of intellect: abstraction of meaning from words, sentences, and paragraphs; addition of fractions; nature study and science; long division; spelling from dictation; and language usage. Among arithmetical processes, increments of intellect are more valuable in dealing with parts than in dealing with wholes; in subtracting more than in adding; and in dividing more than in multiplying.

We know from our tables and graphs that equalization of opportunity does not equalize our groups. There is throughout the three years a marked difference in achievement. It would be of great interest to determine, if possible, the comparative magnitudes of this difference at the beginning and at the close of the period of equal opportunity in special classes. This comparison is not, however, possible by any method known to the present investigators. True comparison in terms of time saved at each extreme is impossible because the scales used are not in terms of equal units from beginning to end; and also because *opportunity is restricted for*

the best children toward the end of their climb by the limitations of the tests of achievement.

Perhaps a word may be added here to call attention to the fact that the achievements in which the two groups differed most are precisely the ones in which opportunity had been most indisputably equal. The world of word, sentence, and paragraph meaning lay open to all alike, as presented, for example, by the special library of the opportunity classes, the public libraries of New York City, and the whole reading environment of the city as it pressed upon the children from every sign board, news stand, and other such agency for the presentation of the printed word.

Conclusions

1. Children of the ages herein studied cannot be equalized in achievement by equalization of opportunity. Those who test higher by available tests of intelligence, are subsequently found to stand higher in achievement both at the outset of a three-year period and throughout the whole three years of equal and special opportunity.

2. A group of children 8 to 11 years old clustering at 165 I.Q. (Stanford Binet) differentiates itself from a group of the same age, clustering at 146 I.Q., in performances covering intervals of time even as brief as 30 seconds.

3. By tests of intelligence at present available *educability* is tested. Comparative achievement under equal opportunity can be foretold from these tests with a high degree of reliability.

4. In the course of experimentation it is found that mental tasks differ widely in the extent to which they utilize increments of intelligence. There are scholastic performances which do not profit from the margin of intellect represented by the difference between 146 I.Q. and 165 I.Q.

5. Apparently, the differences in achievement between the two groups herein compared increase in magnitude as the elements or associations involved in the tasks increase in number; or in other words, as the tasks increase in complexity.

6. Children clustering respectively at + 6 P.E. and + 8 P.E. in mental tests, stand far above unselected children in all performances herein studied, with the exception of penmanship. This superiority of achievement is not to be attributed to the special

opportunity of the former, since those at + 8 P.E. also outstrip those at + 6 P.E., although *special opportunity was equalized for these two groups.*

7. Children clustering at 165 I.Q. do not differ appreciably or consistently from children clustering at 146 I.Q. as respects home ratings (all children were drawn from the public school population of the same city). Nevertheless, there is a wide difference in achievement between two groups thus constituted, under conditions of equal opportunity in the school.

Appendix

Sources of Tests Used in Experimentation

Ayres-Burgess, M., *A Scale for Measuring Ability in Silent Reading.* Russell Sage Foundation, New York City

Buckingham, B. R., *Extension of the Ayres Spelling Scale.* Public School Publishing Co., Bloomington, Ill.

Chapman, J. C., and Cook, S., *Speed of Reading Tests (Grades IV to VIII).* J. B. Lippincott Co., Philadelphia, Pa.

Horn, E., and Ashbaugh, E. J., *Spelling Tests, First Series.* J. B. Lippincott Co., Philadelphia, Pa.

Kelley, T. L., Ruch, G. M., and Terman, L. M., *Stanford Achievement Test.* World Book Co., Yonkers-on-Hudson, New York

Monroe, W. S., *Diagnostic Tests in Arithmetic.* Public School Publishing Co., Bloomington, Ill.

Monroe, W. S., *Standardized Reasoning Test in Arithmetic.* Public School Publishing Co., Bloomington, Ill.

New York Scale for Grading Penmanship, Bureau of Research and Reference, Department of Education, New York City

Terman, L. M., *The Measurement of Intelligence.* Houghton Mifflin Co., Boston, Mass.

Thorndike, E. L., and McCall, W. A., *Reading Scale.* Teachers College, Columbia University, New York City

Thorndike, E. L., *I. E. R. Graded Test of Word Knowledge.* Institute of Educational Research, Teachers College, Columbia University, New York City

Trabue, M. R., *Nassau County Supplement to the Hillegas Scale for Measuring the Quality of English Composition.* Teachers College, Columbia University, New York City

Van Wagenen, M. J., *American History Scales.* Teachers College, Columbia University, New York City

Williams, J. H., *A Guide to the Grading of Homes.* Whittier State School, Calif.

Willing, M. H., *Scale for Measuring Written Composition.* Public School Co., Bloomington, Ill.

PREFATORY NOTE TO CHAPTER II

The data reported by Heilman show beyond reasonable doubt that not over five percent of the enormous variability in school product to be found among children of the same age in ordinary communities (or at least among children of age 10, the age of the children of this study) is due to differences in educational exposure. The application is readily enough drawn that to deny school promotions to children solely because of absence (unless extremely prolonged) is liable to work an injustice upon pupils and to tax school facilities unduly. Regardless of shortage in the educational process of months, or even of a year or two, children tend to attain a proficiency in school subjects close to that appropriate to their mental level. We are reminded of Breed's experiments in which young chicks that were prevented from learning to peck grain during the first few days of life became, when released, as proficient almost at once in the pecking response as control broods which had practiced pecking from the time of hatching! Results such as those of Heilman open the question as to whether eight years of school attendance is really necessary to bring pupils up to the standard usually achieved by the eighth grade. One wonders whether in the four or five years between ten and fourteen they might not learn to read, write, and spell as well, and master as much arithmetic, history, and geography as they would be likely to in eight years.

The author's use of the Wright "path-coefficient method" is of great interest, since this technique has not been applied to educational problems previous to the publication of the present Yearbook. The author's cautiousness in the application of this method, which offers a basis for hypothesis no more sure than does any other technique having less mathematical precision, is especially to be commended.

CHAPTER II

THE RELATIVE INFLUENCE UPON EDUCATIONAL ACHIEVEMENT OF SOME HEREDITARY AND ENVIRONMENTAL FACTORS[1]

J. D. Heilman
Colorado State Teachers College, Greeley, Colorado

Purpose

The specific purpose of this investigation was to determine, in so far as its nature and the development of statistical technique permitted, the relative influence upon scholastic achievement of mental age, school attendance, and socio-economic status of the home, with the hope of throwing some light upon the relative weight of inborn capacity and environmental influences in determining the large individual differences in acquired abilities. In this study acquired abilities are represented by scholastic attainment as determined by the Stanford Achievement Test, environmental influences by school attendance and socio-economic status, and inborn

[1] This investigation was made in response to a request from Dr. L. M. Terman for a contribution to the present Yearbook. I am indebted to Dr. Terman for assistance in planning the investigation, to Miss Barbara Burks for help in the employment of statistical techniques, and to both Dr. Terman and Miss Burks for aid in making an interpretation of the results. The plans for the investigation were completed in October, 1925.

The investigation was made possible through the scientific interest of President G. W. Frasier, Superintendent J. H. Newlon, and Deputy Superintendent A. L. Threlkeld. The entire study was financed by the Colorado State Teachers College and the Denver Public School System at a total cost of $1,000. All of the data were collected in the public schools of Denver under the supervision of Charles E. Greene, the Director of Measurements.

A number of paid assistants were engaged to collect data, administer and score the tests, and tabulate the results. The tests were given by two individuals who had received excellent training and practice for the work. All of the computations were made by me. I wish to express my gratitude to all those who gave assistance in any way in making this investigation.

I also wish to acknowledge my indebtedness to Dr. F. L. Whitney, Director of our Research Department, for permission to use some of the results of an investigation which is now being made by him and his Assistant, Wilford H. Woody. I am further obliged to Miss Henriette K. Woolf for permission to use some of the results of her master's thesis.

In this article only those results will appear which have a direct bearing upon the problems of the present Yearbook. I expect to publish elsewhere additional results which it is hoped will be of practical value to education and its administration. I also intend to publish the re-standardization of the Chapman Socio-Economic Scale which was used in this study.

capacity by mental age as determined by the Stanford Revision of the Binet Scale.

The solution of this problem is of the greatest practical importance to educational and eugenic measures. If native capacities and training, or environmental influences of the school and home, have an equal share in accounting for differences in abilities, then there should probably be a well-balanced emphasis upon educational and vocational guidance on the one hand and training on the other; but the emphasis should be shifted to guidance if inheritance is the important factor and to training if the influence of the environment plays the more important rôle. At present, especially in our high schools and colleges, the factor of guidance is receiving more and more recognition. We are, therefore, in great need of scientific guidance in our educational practice. Similarly, we are unable to deal wisely with the problem of the application of eugenic measures without reliable knowledge concerning the relative influence of nature and nurture.

General Plan

A large group of children, conforming as nearly as possible to the statistical requirements for accuracy, was selected for measurement. For each child included in the investigation the following data were obtained:

1. Educational age in months
2. Mental age in months
3. Life age in months
4. School attendance in days for each grade (and for the kindergarten)
5. Socio-economic status of the home, in terms of S.D.
6. Date of entering the first grade (and of the kindergarten for those children who had kindergarten training)

Educational age was taken as an effect which it was assumed was caused by environmental and hereditary factors. The environmental factors are represented by school attendance and the nature of the home, and the hereditary factors by mental age. No claim is made that all of the environmental influences upon educational age are included in school attendance and the socio-economic status

of the home or that all of the hereditary influences upon educational age are embraced in what is called mental age. Whatever influences have been measured, it is proposed to determine their relative potency in producing variation in educational age under the existing school and social conditions of the communities from which the data were obtained.

This may be done by making use of the fact of variation. We found a very large individual difference in the effect, or educational age—represented by the amount of variation. Similar individual differences, or variations, were obtained for the factors assumed to be causes of the effect. Now, the relative potency of our causal factors in producing educational age can be determined by showing which of them contributes the largest portion to the amount of variation in educational age. We may say that school attendance is the most potent factor if a given amount of change in it produces a more marked change, or variation, in educational age than an equal amount of change in mental age and in socio-economic status will produce. If, on the other hand, variation in educational age is determined more by changes in mental age than by equal changes in either of the other factors, then mental age is the more potent in influencing educational age.

In reading this article it is important to remember that such results as have been obtained do not apply to widely different situations. Larger variations in any one of the causal factors would, no doubt, have increased its weight in contributing to the amount of variation in educational age. Life age for our group of children was restricted to an extreme variation of only nine months. This, undoubtedly, reduced the variation in educational age, mental age, and school attendance. Moreover, the children were selected from schools located in communities of average social standing, and this no doubt reduced the variation of socio-economic status. However, on our measuring instrument for this factor, the highest and lowest possible scores were made.

On account of the very narrow range of life age, the influence of this factor upon the variation of each one of the other factors was eliminated by means of partial correlations. The standard deviation for life age was only 2.6 months, and the coefficients of correlation for life age and educational age and life age and mental

age were only .135 and .120, respectively. No attempt is made, therefore, to determine the relative significance of life age in contributing to educational age.

For the purpose of actually determining the relative weight of the causal factors in producing variation in the effect, the correlation technique was employed.

The technique most commonly employed for the purpose of finding the proportionate contribution of several causes to an effect is the partial and multiple correlation technique. This technique[2] is adequate only if the factors whose true correlation is sought do not contribute to the factors which are rendered constant. But for our factors, or variables, this assumption cannot be justified, because mental age may be expected to contribute to socio-economic status through the resemblance which children bear to their parents in mental age. If this is the case, then the true correlation between educational age and mental age is reduced too much by rendering socio-economic-status constant, because this reduces the size of the coefficient, not only by reducing the variation in educational age, but also, in violation of the principles on which the technique is based, by reducing the variation in mental age. If mental age contributes to school attendance, then the true correlation between educational age and mental age is again reduced too much by rendering school attendance constant.

A technique somewhat better adapted to the solution of our problem is that of path coefficients, suggested to me by Miss Burks. By this method, which will be described more fully in subsequent pages, the direct contributions to the effect made by the different causal factors are also reduced too much, but the reduction, instead of being entirely lost, appears in the joint contribution made by the causal factors to the effect. The method is also far less laborious than the partial and multiple correlation technique.

Our problem presents other difficulties which are probably of minor importance. For example, if the factor taken as the effect and any one of the factors taken as causal have a reciprocal influence upon each other, no correlation technique can determine how much variation in the factor taken as the effect is produced by the

[2] Burks, Barbara S. "On the inadequacy of the partial and multiple correlation technique." *Jour. of Educ. Psych.*, 27: Nov., 1926.

factor taken as the cause. Now, in this investigation, I am of the opinion that grade placement and effort are factors which bear such a reciprocal relation to educational age. Just what their significance is, it is impossible to tell. However, they have been left entirely out of the computations.

It is also possible that school attendance and mental age, socio-economic status and mental age, educational age and school attendance, and educational age and mental age bear a progressive reciprocal relation to one another. Under such conditions the correlations obtained would hold precisely only for children who have reached a stage in their chronology represented by the age of the subjects of this study (ten years). Moreover, in the event of reciprocal action between educational age and any of the other variables, we could conclude only that the relations uncovered by the statistical treatment employed indicated a probable *upper limit* which the *true causative effects* of certain variables upon the criterion would not exceed. For example, later on, where the statement is made that about 45 percent of the variability in educational age is due to mental age, about five percent to school attendance, and about nine percent to the combined effects of mental age and school attendance, a more cautious interpretation might waive the idea of causation and state merely that mental age and educational age have a considerable degree of influence upon one another, and that educational age and school attendance contain only a small degree of mutual implication. The *direction of influence* may be partly or even largely from educational age to mental age and school attendance, rather than from school attendance and mental age to educational age, though the latter seems the more probable. But it is certainly significant to know that if differences in school attendance *could be equalized* experimentally without simultaneously altering the composition of the group as regards age, mental age, and socio-economic status, the variance of the group in educational age would not be decreased by more than thirteen percent, and possibly by less.

Another point in which caution of interpretation must be used is connected with the assumption that the variables whose 'effect' upon achievement we seek to measure account *in themselves* for the influences found. No one could prove without a fundamental and

carefully controlled experiment whether or not an apparent 'unique' contribution of socio-economic status to educational age was due actually to environment or to definitely heritable, but unmeasured factors, other than Binet intelligence, *which are correlated with socio-economic status.* Similarly, no one could prove that the apparent contribution of Binet mental age was due entirely to intelligence, or in part to unmeasured factors, such as interest or effort, which are probably correlated with intelligence. In view of such considerations, it should be doubly emphasized that the relations found in this study represent only probable *upper limits* beyond which the true influences of the measured 'causes' do not lie.

The Data

The subjects of the investigation were 828 school children who, on March 1st, 1926, ranged in age from 10 years, 1 month, to 10 years, 10 months. All but 50 of these children had attended no other than the Denver schools. These 50 had not attended any other schools more than two years. All children who had foreign parentage and whose school records were incomplete were excluded from the investigation. All other 10-year-olds in each of the 48 schools involved were included in the investigation. It is believed that this is an unselected group. Thomson and Pintner[3] say "one criterion for the unselected nature of any group is that the coefficient of correlation of age and mental age in such a group should be approximately equal to the ratio of the coefficients of variability of chronological and mental ages." For this group the coefficient of correlation of age and mental age is .120, and the ratio of the coefficient of variability of age to that of mental age is .142. These figures, accordingly, meet the requirement of approximate equality.

Practically all of the data were collected by the two examiners already referred to, Mrs. Gladys May Maclin and Mrs. Dallas Allhands. They began their work on January 25, 1926. On the same day Superintendent Threlkeld sent a letter to all of the principals and directors of the schools involved in the investigation, in which he explained the nature and purpose of the investigation

---

[3] Thomson, Godfrey H., and Pintner, Rudolph. "Spurious correlation and relationship between tests." *Jour. of Educ. Psych.*, 15: October, 1924.

and asked them to coöperate. As outlined by Mr. Greene, the Director of Measurements, the examiners first listed the names of all the children who were born between May 1, 1915, and January 30, 1916. They next determined which of these children had attended no other than the Denver schools. The next step was to send questionnaires to the parents to secure information in regard to the homes. This was followed by giving the achievement tests. The remainder of the time was devoted to the giving of mental tests.

All of the age and attendance data were taken from the records of the Denver schools. The occupation of the father was also taken from the records. These were copied on blanks especially prepared for the purpose, of which the following is a sample:

BLANK FOR AGE AND ATTENDANCE DATA

Fill out this blank only for children whose birth dates lie between May 1, 1915, and January 30, 1916, inclusive. Do not include children of foreign parentage and children whose records are incomplete.

Name of child..................Birth date............Date.....

School ....Grade..................

Date of entering the kindergarten.............................

Date of entering the first grade..............................

List the attendance in days in each of the following:

Kindergarten ....Grade III..........Grade VI............

Grade I ....Grade IV...........Grade VII...........

Grade II ....Grade V............Grade VIII..........

Occupation of father..

The educational age was based upon the results of the Stanford Achievement Tests. These tests were given during the two weeks preceding and the two weeks following March 1, 1926. As the probable error on this test for ten-year-olds is 1.5 months, no correction was made for differences in the time of administration. All of these tests were scored and the results tabulated by experienced students of the Colorado State Teachers College under the supervision of an assistant in the Department of Psychology.

The mental age of the children was determined by means of the Stanford Revision of the Binet Scale. This test was selected in preference to group tests because it is practically free from the influence of school training. The tests were given from the middle of March to the end of the school term about the middle of June. By means of the I.Q. the mental ages of the children were all corrected to March 1, 1926. Both assistants who administered the tests were experienced in Binet procedure. Mrs. Maclin received part of her training in giving Binet Tests from Dr. Terman and part of her training from Mr. Greene in a summer course at the University of Denver. Mrs. Allhands received her training at the University of Denver under Dr. Garth.

The method of checking on the tests was as follows: A small number of tests, from six to twelve, were first given by each examiner, then checked by both; later, they were checked by Miss Florence Barney, who does most of the Binet testing in the Denver schools and who is well trained and very capable in this work. Each week of the eighteen Mrs. Maclin and Mrs. Allhands checked their method of scoring and administering the Binets.

The attendance of the school children was determined up to the end of the first semester, January 25. As the achievement tests were given around the first of March, no large discrepancies in attendance could have occurred between these two dates. Large errors in attendance might presumably have affected the scores on the achievement tests.

The cultural and economic data, exclusive of the father's occupation, were obtained from the home by means of a questionnaire. The data and the method of securing them are sufficiently explained by the following letter and "Question Blank for Parents." If Item 12 of the blank—"Do you live in an apartment house?"—was answered in the affirmative, Item 13 concerning the number of rooms in the house was not used in determining the score.

TO PARENTS:

The National Society for the Study of Education has asked the Denver Public Schools to coöperate in making a study of normal ten-year-old children to discover the importance of regular attendance. Most of the necessary information can be secured from school records. The information which is asked for below may seem to be of a very personal nature, but it has a bearing on educational results. It will be

used only for this study and by those who are interested in the results from the city as a whole. It will not affect the school standing of your child and will not be seen by the teachers. The forms will be destroyed as soon as the results are compiled.

It is important that the blanks be filled out accurately and completely. We appreciate your coöperation.

CHARLES E. GREENE,
Director of Measurements
Denver Public Schools

DENVER PUBLIC SCHOOLS

AN INVESTIGATION RELATING TO POSSIBILITIES AND LIMITATIONS OF TRAINING

QUESTION BLANK FOR PARENTS

Name of child........................Date........School..........

1. Do you have a telephone in your home?........................
2. Did the father of the child attend high school?..................
3. Did the mother of the child attend high school?................
4. Is English the only language spoken in the home?...............
5. Is your house heated by a furnace?...........................
6. Do you have an auto other than a truck?.......................
7. Do you have a piano in your home?...........................
8. Do you have a phonograph in your home?......................
9. How many daily papers do you take regularly in your home?......
10. How many magazines do you take regularly in your home?........
11. About how many books are there in your home?..................
12. Do you live in an apartment house?..........................
13. How many rooms, not including bathroom, are there in your home?.............

....
(Name of parent)

Note—Please put this question blank into the attached envelope. Seal the envelope and ask your child to hand it to his teacher to-morrow morning.

The information asked for on these blanks was complete for 688 cases. The data on all of these cases were used for the purpose of making a socio-economic scale. The technique used in making the scale differed only in minor particulars from that used by Chapman and Sims.[4] Their scale was not used because we included two items, number of rooms and occupation, which do not appear in their scale. Moreover, the item on the number of magazines had to be treated differently to make its inclusion significant. The scores on each of the items are expressed in sigma times ten. The lowest score that can be made on the entire scale is 578, and the highest 702. As the information was incomplete for 140 cases, owing chiefly to the decease of the father and residence in apartment houses, the total scores made on the 13 items were averaged and the result was multiplied by 10. This reduced the lowest possible score to 445 and the highest possible score to 540.

The reliability of the scale was determined by computing a coefficient of correlation between the halves of the scale. The scores made on Items 1, 2, 8, 9, 11, and 13 formed one half, and the scores made on the remaining items the other half. The attempt was made to pair the items in such a way that those in the one half would represent as nearly as possible the same things as the items in the other half. The coefficient for the two halves is .77; that between the results of this scale and those of a similar one, .87. Corresponding figures for the Chapman-Sims scale are .63 and .77, respectively. The coefficient of reliability for the revised scale is somewhat higher than it should be because about 50 of the 688 children made perfect scores.

The schools involved in the investigation were selected as nearly as possible in districts of average economic status. This was done to equalize as much as possible the socio-economic factor. However, the results of the scale show that every grade of socio-economic status provided for by the scale was represented in all of the schools combined.

The investigation by Miss Woolf[5] was patterned after this one and is almost an exact duplicate of it. The same tests were used. The

---

[4] Chapman, J. C., and Sims, V. M. "The quantitative measurement of certain aspects of socio-economic status." *Jour. of Educ. Psych.*, 26: September, 1925.

[5] Woolf, Henriette K. *The Relative Influence of Some Environmental Factors and Mental Age on Educational Age.* Unpublished Master's Thesis, August, 1926. Colorado Teachers College, Greeley, Colo.

Chapman-Sims[6] Scale was changed so as to include "number of rooms" and "the father's occupation" and was then re-standardized with the Chapman-Sims technique. The subjects were all of the 10-year-old children of the Greeley Schools. Computed to March 1, 1926, the children ranged in age from 9 years, 9 months, to 10 years, 9 months, with an average age of 10 years, 2 months. Miss Woolf did all of the work personally, for which she was well prepared, having taken her A.B. degree at Stanford University, where she majored in the Department of Psychology. Perhaps the most significant difference between the two investigations is the difference in the number of cases—her cases numbered only 157. In interpreting her results she employed the partial coefficient and regression equation technique. However, she used the score-form of the equation and weighted the coefficients with the size of the arithmetic means. I changed this form of the equation into the special deviation form with the sigmas equal, and computed a similar equation with my data in order to make the results of our investigations comparable.

Dr. Whitney is at present engaged in making a very comprehensive study of "The Selection, Achievement, and Advancement of Secondary School Children." In making this investigation he used the Stanford Achievement Scale, the Multi-Mental Scale and the Chapman-Sims Scale to test a very large sample of children of the state of Colorado. In addition to these measurements, he obtained data on school training (number of years in school), distance from school, physique (departure from normal weight for a given height and age), permanence of residence, and life age.

In connection with the general investigation, these data were obtained for the ninth-grade children in the over-three-teacher schools of Weld County, Colorado. In the statistical treatment of these data, partial correlations and the regression equation in deviation form with sigmas constant were employed. For the ninth-grade girls the method of path coefficients was also used. Educational age was taken as the dependent variable. As the data were not complete for all of the children, the zero-order correlations for the boys involved from 233 to 292 cases, and for the girls from 228 to 308 cases.

The Whitney-Woody investigation, however, differs from this one and Miss Woolf's in important respects. The children were much older and differed far more widely in life age—the coefficient of variation for their group of girls was about 3 times the size of the coefficient of variation for the group of children of this investigation. The basis of selecting their children was that of membership in the ninth grade in the over-three-teacher schools of Weld County. More factors were included in the list of independent variables and mental age was measured by means of a group test. The attendance was only roughly determined by having the teachers report the number of years each child attended

[6] Chapman-Sims. *Loc. cit.*, page 13.

school. In case a child attended different schools, the teacher of the school in which the child was located at the time of the investigation wrote for the child's attendance records in the schools previously attended. Moreover, the majority of the children were rural, belonging to the farming class. It is, therefore, to be expected that their results will agree less closely than Miss Woolf's with those of this investigation.

Results

1. Effect of Kindergarten Training

On account of the somewhat prevalent belief that early training is far more efficacious than later training, I have compared the scholastic attainments of the kindergarten with the non-kindergarten trained children of our group. Of the 828 ten-year-olds, 205 had no kindergarten training, 615 had attended the kindergarten from 20 to 185 days with a median attendance of 132.5 days, and 8 had attended fewer than 20 days. The records of the 8 children who had attended fewer than 20 days were not included in the computations.

In Table I a comparison is made on the basis of the average of the EA, MA, SA, and SES of the kindergarten and non-kindergarten groups.

Table I.—Average Educational Age, Mental Age, School Attendance, and Socio-Economic Status of Kindergarten and Non-Kindergarten Groups

	Aver. E. A. in months	Aver. M. A. in months	Aver. S. A. in days	Aver. S. E. S. in sigma
Kindergarten.........	139.69	134.39	701.48	502.07
Non-Kindergarten.....	137.56	132.67	655.98	494.63
Difference..........	2.13	1.72	45.50	7.44

The kindergarten group surpasses the non-kindergarten group by two months in educational age, but in mental age it also excels and by almost the same amount. Moreover, in school attendance and socio-economic status the kindergarten group excels by 45.50 and 7.44 points, respectively. In order to determine the significance of these differences, I computed the S. D. of the difference for each of the four items. These are given in the following tabulations:

TABLE II.—S. D. OF THE DIFFERENCES BETWEEN THE KINDERGARTEN AND NON-KINDERGARTEN GROUPS REPORTED IN TABLE I

	S.D. (dis)	S.D. (aver)	S.D. (diff)	Diff. ÷ S.D. (diff)	Chances in 100 that Diff. is Greater than 0
Educational Age					
Kindergarten.......	14.69	.592	1.27	1.68	96
Non-Kindergarten..	16.17	1.129	....	....	....
Mental Age					
Kindergarten.......	19.36	.781	1.59	1.08	86
Non-Kindergarten..	19.78	1.381			
School Attendance					
Kindergarten.......	98.99	3.992	9.09	5.01	100
Non-Kindergarten..	116.94	8.167			
Socio-Economic Status					
Kindergarten.......	27.61	1.113	2.22	3.35	100
Non-Kindergarten..	27.50	1.921			

If the differences between the averages given in Table I are to be considered significant differences, they should be at least three times the size of the S. D. of the differences. As shown in the fourth column of Table II, this is not the case for either educational age or mental age, but it is the case for both school attendance and socio-economic status. There is, therefore, no significant difference between the two groups in either educational age or mental age, in spite of the fact that there is a significant difference between the groups in school attendance and socio-economic status, in which the kindergarten group excels. Despite the fact that the kindergarten group attends school better and has a better socio-economic status, it does not give evidence of a real superiority in either educational age or mental age.

Another method of making these comparisons is to compute, by means of the bi-serial r method, coefficients of correlation between kindergarten attendance and each one of the four variables. The results of these computations are set forth in the following tabulation:

	Coefficients	Standard Error
KA-EA	.0834	.0498
KA-MA	.0521	.0500
KA-SA	.2517	.0478
KA-SES	.1583	.0492

All of these coefficients are positive, but only in the cases of KA-SA and KA-SES are the coefficients more than three times the standard errors. It is only in these two cases that we can be reasonably sure that a true positive correlation is indicated by the obtained correlations. These results, then, are in accord with the conclusions just expressed with reference to the effect of kindergarten attendance upon the scholastic achievement of the ten-year-olds.

It might be assumed that the differences between the two groups in school attendance and socio-economic status were not sufficient to have any significant effect upon school achievement. If we grant this assumption, and express the achievement only with reference to the average mental ages of the two groups by computing the accomplishment ratios from the average E.A.'s and M.A.'s, we obtain for the kindergarten group an A.R. of 103.9 and for the non-kindergarten group an A.R. of 103.7.

From the three different methods of treating the data, no evidence appears that ten-year-old children who have a median kindergarten attendance of 132.5 days will do any better on the Stanford Achievement Test than the ten-year-olds who have never attended the kindergarten. However, these results should not be interpreted to mean that kindergarten training is of no importance, for such training may adapt the children better to their school life in the first grade, and aid the children in acquiring many other responses not usually regarded as scholastic in character.

2. Results of Measurements

As previously stated, the children varied only 9 months in life age, but in spite of this narrow range in life-age they varied considerably in the other measurements. In Table III are given the measures of variability and central tendency for educational age, mental age, school attendance, socio-economic status, life-age, and grade placement.

TABLE III.—EXTREME AND TENTH TO NINETIETH PERCENTILE POINTS

	Extreme Points	Difference	10th to 90th Percentile	Difference
E. A. in months....	90—205	115	121.61—158.28	36.67
M. A. in months....	80—215	135	110.14—158.52	48.38
S. A. in days........	300—1000	700	541.80—795.61	253.81
S. E. S. in S. D.....	445—545	100	459.48—536.82	77.34
L. A. in months....	121—130	9	121.99—129.13	7.14
G. P. in grades......	2— 6	4	3.55— 5.36	1.81

The differences in the first column of differences are somewhat larger than they should be for E.A., M.A., and S.E.S., because they represent the lower and upper limits of the lowest and highest intervals of the distribution tables. Five points were included in each interval. All ranges of these three items are also somewhat too large, because the reliability of the measuring instruments was not perfect. However, for E.A. and M.A. the reliability of the tests is so high as to make the error for these items negligible.

The distributions, excepting those for L.A. and S.E.S., conform fairly well to the requirements of a normal distribution. For L.A. and S.E.S. the distributions are more nearly rectangular. The manner of selecting the children explains this type of distribution for L.A. However, on account of the fact that the ages were computed only to the nearest month, the first and last intervals of the L.A. distribution have only about one-half as many cases as the other intervals. The rectangular nature of the distribution for S.E.S. may be partly due to the fact that the children were selected only from those schools which were located in communities of average socio-economic status. As the G.P. was determined for the children during the first semester and the tests were given during the second semester, the extreme points for G.P. at the time the tests were given would be 2½ and 6½, if the extreme cases received a promotion at the end of the first semester.

Children who were in the lowest division of Grade II received a score of 2, while those in the highest division received a score of 2½ for G.P. Similar scores were given to children in the other grades.

If the 10 percent of the children who made the lowest scores and the 10 percent who made the highest scores are eliminated from the computations, the range for the remaining 80 percent is still very

large for a group of children whose extreme range in life age is not more than 9 months, as may be seen in the last column of the table. These remaining 80 percent have an extreme range of 3 years in E.A., 4 years in M.A., 254 days in S.A., and 2 grades in G.P.

Additional measures of variability, as well as measures of central tendency, are given in Table IIIA.

TABLE IIIA.—MEASURES OF CENTRAL TENDENCY AND VARIABILITY

	Arithmetic Mean	Standard Deviation	Coefficient of Variability	Coefficient of Variability (Woolf)
E. A. in months.......	139.24	15.06	.1082	.1428
M. A. in months......	134.05	19.51	.1455	.1353
S. A. in days.........	690.32	104.30	.1511	.1698
S. E. S. in S. D.......	500.29	27.73	.0554	.0740
L. A. in months.......	125.78	2.607	.0207	.0282
G. P. in grades.......	4.28	.656	.1533	.1733

The average attendance is given only to the end of the first semester, 1926. The average G.P. is that which obtained for the first semester. Therefore, for the second semester, if practically all of the children were promoted, the average G.P. should be 4.78, or .5 of a grade higher. As the units of the different measuring instruments which were used in the investigation differed in size, the coefficient of variability affords a better basis for comparing the variability of the different variables than does the standard deviation. These values are given in the third column of Table IIIA. Upon the basis of these coefficients, M.A., S.A., and G.P. are approximately equal in variability. For E.A., variability is somewhat lower. It is still lower for S.E.S., and, of course, considerably lower for L.A.

In the last column of Table IIIA, the coefficients of variability for Miss Woolf's investigation are given. Excepting M.A., all of her coefficients are somewhat larger, as they should be, because for her children the extreme range in life age was 3 months more than for the children of this investigation. S.E.S. is larger because no attempt was made to select schools in communities of average social status, as was the case in this investigation. In spite of this fact, there is practically no difference between the two coefficients for S.E.S.

In Tables III and IIIA it has been shown that there is a large amount of variability in E.A. The important fact to be determined in this investigation is the amount of variability in E.A. which may be ascribed to each of the factors which have been selected as the causes of this variability. If the major portion of the variability in E.A. is due to S.E.S. and S.A., then we would have to conclude that the environmental forces as measured by these factors were the most potent in determining differences in educational results as measured by E.A. If, on the other hand, the major portion of the variability in E.A. is due to M.A., then we would be justified in concluding that heredity, as measured by M.A., was most potent in determining differences in education. If, in the third place, a large portion of the variability in E.A. remains unaccounted for, it will be impossible to determine definitely and conclusively from the data of this investigation whether heredity or the environment is the main cause of the differences in the scholastic abilities of our group of children under the present environmental circumstances of school and social life.

In order to determine which of our several factors contributes the most to the variability in E.A., the correlation technique has been employed. In Table IV the zero-order coefficients are given for the different factors. The factor of grade placement does not appear in this table because, as previously pointed out under the general plan of the investigation, our technique is inadequate to deal with this particular influence.

As it appeared from an inspection of the correlation tables that all correlations were of the linear type, all zero-order coefficients were computed by using the usual Pearson formula.

TABLE IV.—ZERO-ORDER COEFFICIENTS

	M. A.	P. E.	S. A.	P. E.	S. E. S.	P. E.	L. A.	P. E.
E. A.	.737	.0106	.447	.0188	.356	.0205	.135	.0230
M. A.	....	.....	.317	.0211	.359	.0204	.120	.0231
S. A.	....	.....	....	.....	.169	.0228	.415	.0194
S. E. S.	....	.....	....	.....	....	.....	.001	.0217

The coefficient for L.A. and S.E.S. is insignificant, as it should be expected to be. The coefficients for L.A. and E.A. and for L.A. and M.A. are only slightly significant. These, as well as the coefficient for L.A. and S.A., would have been much higher had the chil-

dren not been so selected as to limit the range in life age to nine months.

It might also be assumed that, as an attempt was made to select schools in communities of average social status, the coefficients for S.E.S. and each of the other variables are too low. This may be the case, but Miss Woolf[7] who included in her investigation all of the ten-year-olds of the Greeley Schools, exclusive of those for foreign-speaking parentage, obtained even lower correlations for S.E.S. and the other variables. I am giving her correlations in the following tabulation, excepting the one for S.E.S. and L.A., which she did not compute.

S.E.S.-E.A., .369
S.E.S.-M.A., .257
S.E.S.-S.A., .056

The coefficient for E.A. and S.A. is lower than it should be if, as some contend, absence from school does not have any significant effect until a certain amount has been reached. While the number of cases with different amounts of absence is probably too few to furnish any very reliable evidence on this point, I am, nevertheless, giving in Table V medians in E.A., M.A., and S.E.S. for different amounts of school attendance.

TABLE V.—MEDIANS IN EDUCATIONAL AGE, MENTAL AGE, AND SOCIO-ECONOMIC STATUS FOR DIFFERENT AMOUNTS OF SCHOOL ATTENDANCE

Attendance in Days	Med. E. A. in Months	Med. M. A. in Months	Med. S. E. S. in S. D. Units	Frequencies
820–900	151.7	143.3	515.0	38
780–	*141.8	*134.2	*506.4	98
740–	141.7	137.1	505.1	169
700–	140.8	*134.8	507.8	119
660–	*135.6	132.8	*502.8	121
620–	133.5	134.1	501.6	70
580–	*129.8	*122.9	*489.7	81
540–	132.5	126.5	498.8	43
500–	*127.5	*123.1	*483.8	33
400–	127.1	124.0	*478.3	40

[7] Woolf, Henriette K. *Loc. cit.*, p. 14.

The attendance of each of the ten groups (Table V) is less by 40 days than that of the preceding group, excepting the last group which attended 100 days less than the preceding group. Asterisks indicate the largest group differences in E.A., M.A., and S.E.S. The most marked decrease in E.A. occurs between the groups of best and next best attendance. However, in consideration of the fact that the attendance range of the best attendance group is twice that of the other groups, the decrease should be reduced by one-half. Even with this reduction, it is as large as any of the decreases. The data do not, therefore, show that a decrease in E.A. is more likely to be associated with a decrease in attendance in one than in any other part of the entire range of attendance.

It appears, moreover, from a comparison of the largest differences in E.A. with the largest differences in M.A. and S.E.S. as indicated by the asterisks, that they occur with surprising regularity between the same groups. The connections among these variables appear to be far more intimate than between any of them and school attendance.

On account of variable or observational errors in measurement, all of the coefficients involving E.A., M.A., and S.E.S. are too low. To correct for this attenuation, I used the following formula:

$$r_{AB} = \frac{r_{A_1B_1}}{\sqrt{r_{A_1A_2} r_{B_1B_2}}}$$

In this formula r_{AB} represents the true correlation between any two tests, A and B; $r_{A_1B_1}$ represents an obtained correlation between any two tests A_1B_1, and the expressions in the denominator represent coefficients of reliability for any two tests.

The coefficients involving E.A., M.A., and S.E.S., corrected for attenuation, are as follows:

$r_{EA-MA} = .7709$	$r_{MA-SA} = .3287$	$r_{SA-SES} = .1812$
$r_{EA-SA} = .4497$	$r_{MA-SES} = .3991$	
$r_{EA-SES} = .3840$	$r_{MA-LA} = .1244$	
$r_{EA-LA} = .1358$		

The self-correlations used in making the above computations are:[8]

$R_{EA} = .988$ $R_{MA} = .93$ $R_{SA} = 1.00$ $R_{SES} = .87$ $R_{LA} = 1.00$

[8] For the first of these, see Kelley, T. L., Ruch, G. M., and Terman, L. M. *Stanford Achievement Test. Manual of Directions.* World Book Co., 1924; for the second, Terman, L. M., and DeVoss, J. C. "The educational achievements of gifted children." *The Twenty-Third Yearbook* of this Society, Part I, 1924.

As the subjects of this investigation ranged in age only 9 months, and were located in schools selected upon the basis of average social standing, the reliability coefficients for E.A., M.A., and S.E.S. might have been increased somewhat if they had been corrected for range, but as this was not possible for S.E.S. (on account of the absence of the necessary S.D.), the correction was not made for any of them.

On account of the narrow range of L.A., which has the effect of reducing the sizes of the coefficients between it and the other variables, it would give a false picture of the relative influence of the different independent variables upon variability in E.A. if its influence were included along with the influences of the other variables. Therefore, the influence of L.A. was rendered constant by partialling it out from the coefficients as corrected for attenuation. The corrected coefficients were modified by this procedure and are set forth in the following tabulation:

$r_{EA\text{-}MA} = .7670$	$r_{MA\text{-}SA} = .3070$	$r_{SA\text{-}SES} = .1987$
$r_{EA\text{-}SA} = .4363$	$r_{MA\text{-}SES} = .4021$	
$r_{EA\text{-}SES} = .3875$		

These are the coefficients which were used to determine the relative importance of M.A., S.A., and S.E.S. in producing variability in E.A., both by means of the regression equation and the method of path coefficients. However, neither of these methods establishes a causal relationship between the dependent variable on the one hand and the independent variables on the other. Such a relationship must be assumed, or shown in other ways. These methods simply show how much variability in E.A. is associated with a unit's variability in each of the independent variables. If reason leads us to accept a causal connection, then the relative influence of M.A., S.A., and S.E.S. upon E.A. follows.

The special form of the regression equation with S.D. for each of the variables equal to 1 was employed. Here are both the formula and the equation.

$$X_{EA} = b_{EA,\ MA.\ SA,\ SES}X_{MA} + b_{EA,\ SA.\ MA,\ SES}X_{SA} + b_{EA,\ SES.\ MA,\ SA}X_{SES}$$
$$X_{EA} = .6711X_{MA} + .2155X_{SA} + .0751X_{SES}$$

As a check on the accuracy of these computations, the coefficient of multiple correlation was computed by two different formulae. The following multiple coefficients were obtained:

$$R_{SA\ (MA,\ SA,\ SES)} = .7985$$
$$R_{SA\ (MA,\ SA,\ SES)} = .7982$$

If, as stated before, a causal relationship is assumed between E.A. and each of the independent variables, then a change of 1 unit in M.A. will produce a change or variation of $\frac{2}{3}$ of a unit in E.A.; a variation of one unit in S.A. will produce a variation of only $\frac{1}{5}$ of a unit in E.A.; and a variation of 1 unit in S.E.S. will produce a change of less than $\frac{1}{10}$ of a unit in E.A.

In order to bring my results into as intimate relation as possible with those found by Miss Woolf,[9] I computed a regression equation for these four variables without correcting for attenuation and without rendering life age constant. These computations were not made by Miss Woolf. The same formula was employed for each of the equations, but the computations were based upon the coefficients found in Table IV. I again checked my computations by means of the method of multiple correlations, obtaining .7767 by one formula and .7772 by the other.

Comparable Results of Other Investigations

The comparable equations obtained by Miss Woolf and by me are as follows:

Woolf . . . . $X_{EA} = .6285X_{MA} + .1734X_{SA} + .1989X_{SES}$
Heilman . . . . . . $X_{EA} = .6366X_{MA} + .2311X_{SA} + .0988X_{SES}$

The correspondence between these equations appears to me to be surprisingly close when it is remembered that Miss Woolf had only 157 cases with wider variations in L.A. and S.E.S., and that her investigation was made in a small town of about 12,000 inhabitants, while the data for this investigation were obtained in a city of about 300,000 inhabitants. Unfortunately, Miss Woolf does not give any check on the accuracy of her computations.

Dr. Whitney and Mr. Woody express their results in terms of the regression equation. These results are subject to the limitations of the partial correlation technique which have been mentioned. Whitney and Woody's investigation differed materially from this one in the number of children and their age, the method of selecting them, the nature of the community, and the number of variables. Moreover, different tests were used for the purpose of measuring intelligence and socio-economic status. It should be noted especially that the children differed widely in life age, as an unselected group of ninth-grade children would. The variables with their coefficients and the squares of the coefficients are given in the following tabulation:

[9] Woolf, Henriette K., *Loc. cit.*, p. 14.

	Regression Coefficients		Squares of Coefficients	
	Boys	Girls	Boys	Girls
1. Mental age............	.4416	.5038	.1950	.2538
2. Life age..............	−.2572	−.2544	.0662	.0647
3. Socio-economic status ..	.1069	.0080	.0114	.0001
4. School training.........	−.0752	−.1242	.0057	.0154
5. Distance from school...	−.0457	.0155	.0021	.0002
6. Physique..............	−.0321	−.0810	.0009	.0066
7. Permanence of residence	.0040	.0786	.0000	.0062

The approximate percentage contribution of each of the causal factors to the effect, in this case educational age, may be read from the last two columns of the tabulation. There are only two significant causal factors, mental age and life age. However, all of the figures are undoubtedly too low on account of the technique employed. As the coefficient for life age is negative, age depresses achievement, as was to be expected for a group of ninth-grade children, the older and duller children being classed with many younger and brighter children. Mental age is, therefor, the only positive coefficient whose contribution to variation in educational age is of any significance.

Use of the Method of Path Coefficients

At the suggestion of Barbara Burks, and with her assistance, I applied to my data the method of path coefficients described by Sewell Wright.[10] This method has several advantages over the ordinary correlation method. In a complex system of correlated variables, the coefficient of correlation gives the total amount of relationship between two variables due to all of the paths by which they are connected, combined, while the path coefficient measures the direct influence of each variable upon any other with which it is connected, independently of the effect of other variables. For example, in our problem M.A. has a direct effect upon E.A., but it also has indirect effects upon E.A. through S.E.S. and S.A. The coefficient of correlation represents the combined direct and indirect effect of M.A. upon E.A., but the path coefficients give the direct effect and all of the indirect effects separately.

However, whether one variable has a *causal* effect upon another must, by this method as well as by all correlation methods, be assumed or determined in other ways.

The symbol for the path coefficient is $P_{X.A}$, in which X is the dependent variable or the effect, assuming a causal relationship,

[10] Wright, Sewell. "Correlation and causation." *Jour. of Agricultural Research*, 20: No. 7, 1921.

and A is the independent variable, or cause. The path coefficient is defined as the ratio of the standard deviation of X due to A to the total standard deviation of X. It is, therefore, the coefficient for the path of influence from A to X.

Another coefficient discussed by Wright [10] is the "coefficient of determination," represented by the symbol $d_{X.A}$ which is equal to $P^2_{X.A}$. This measures the percent of influence which A has upon X in a given system of correlated variables. If all of the causes are accounted for, the sum of the coefficients of determination equals 1.

In Fig. I the variables of our problem, with their path coefficients, are represented.[11]

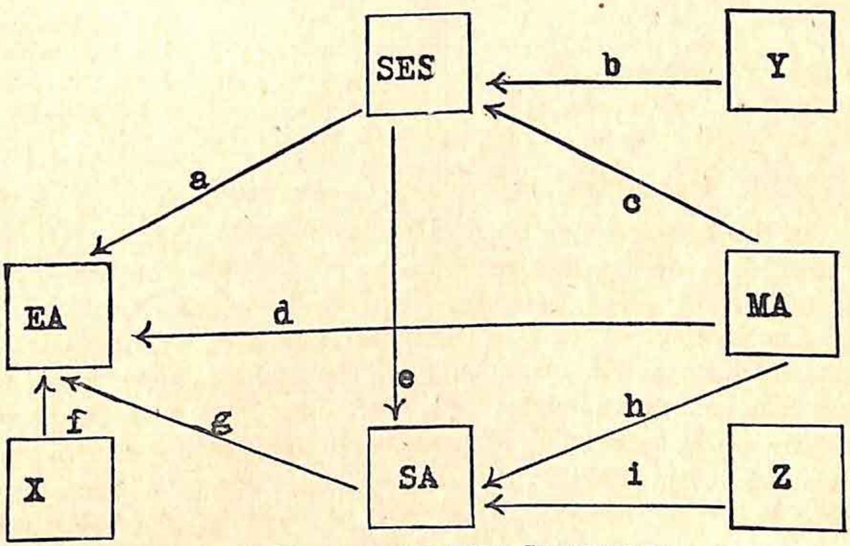


FIG. I. VARIABLES AND PATH COEFFICIENTS

In this diagram:

EA = educational age
MA = mental age
SA = school attendance
SES = socio-economic status
X = complex of all factors, other than MA, SA, and SES, determining EA
Y = complex of all factors, other than MA (or more accurately, other than that portion of heredity which correlates perfectly with MA), determining SES
Z = complex of all factors, other than MA and SES, determining SA

[11] This diagram, with the equations for computing the path coefficients, was furnished by Miss Burks.

The lines in the diagram represent the path coefficients. They represent not only the amount of direct influence of one variable upon another, but also the direction of the influence. These paths must be set up upon the basis of deductions regarding the influence of one factor upon another. As has been said, we have no sure knowledge of whether the directions of influence are actually as they have been represented, whether some of these directions may be reversed, or whether *reciprocal* influences may not account for some of the relationships. The directions of paths which are here postulated seem to me the most reasonable of possible ones on *a priori* grounds. However, the correlations between S.E.S., M.A., and S.A. might be due to an entirely different system of influences than that postulated. But even if this should be the case, the computed path coefficient of *each with E.A.*, upon which the chief conclusions of this study are based, would remain unchanged. In the event of reciprocal action between E.A. and any of the three variables S.E.S., M.A., or S.A., the path coefficients would indicate an *upper limit* which the true causative effect of these variables probably would not exceed. The paths are labeled with small letters. What each of them represents may be gathered from the following equations:

$$
\begin{array}{lll}
P_{EA.MA} = d & P_{SA.MA} = h & P_{SES.MA} = c \\
P_{EA.SA} = g & P_{SA.SES} = e & P_{SES.Y} = b \\
P_{EA.SES} = a & P_{SA.Z} = i & \\
P_{EA.X} = f & &
\end{array}
$$

Because the coefficient of correlation between any two variables is equal to the sum of the products of the paths in each series of paths with which they are connected, plus the path of direct connection, it is possible upon this basis, if the coefficients of correlation are known, to set up enough simultaneous equations to determine the value of each path represented by the letters. This will give the following equations:

$$
\begin{array}{ll}
r_{EA\text{-}MA} = ac + gh + d + ceg & r_{MA\text{-}SES} = c \\
r_{EA\text{-}SES} = a + eg + cgh + cd & r_{MA\text{-}SA} = h + ce \\
r_{EA\text{-}SA} = g + ae + ach + dh + cde & r_{SES\text{-}SA} = e + ch
\end{array}
$$

In these equations all of the coefficients of correlation as well as the value of c are known. From these data the value of each path coefficient can be determined. They are as follows:

$$
\begin{array}{lll}
a = .0750 & e = .0898 & b = .9156 \\
d = .6706 & c = .4021 & i = .9481 \\
g = .2155 & h = .2709 & f = .6022
\end{array}
$$

By a comparison of the path coefficients a, d, and g with the coefficients of the regression equation computed with the same zero order coefficients, I found that they were equivalent. Wright's path coefficient is, in this situation, therefore, the same as the coefficient of the regression equation. However, his coefficient of determination is equal to the square of the path coefficients or the square of the coefficients of the regression equation. Now, as Wright[12] has shown that, where the relation is causal, the coefficient of determination measures the percent of influence which any one of a number of causes has upon their combined effect, or the percent of influence which each independent variable has upon the variability of the dependent variable, it follows that the squares of the regression coefficients measure the percent of influence of each independent variable upon the dependent variable. It also follows that the relative weights of the regression coefficients for determining relative influence are not the ratios of these coefficients, but the ratios of their squares.

Wright[13] says: "It can be shown that the squares of the path coefficients measure the degree of determination by each cause. If the causes are independent of each other, the sum of the squared path coefficients is unity. If the causes are correlated, terms representing joint determination must be recognized. . . . The squared path coefficients and the expressions for joint determination measure the portion of the squared standard deviation of the effect due to the causes singly and conjointly, respectively."

Therefore, in order to determine the fractional part which each of our independent variables contributes to their combined effect or to the amount of variation of the dependent variable, it is necessary to compute the coefficients of determination.

The fundamental equations for these coefficients are:

$$d_{X.A} = P^2_{X.A.}$$
$$d_{X.\overline{AB}} = 2P_{X.A}P_{X.B}r_{AB}$$
$$\Sigma d_{X.A} + \Sigma d_{X.\overline{AB}} = 1$$

---

[12] Wright, Sewall. *Loc. cit.*, page 31.

[13] Wright, Sewall. "Systems of mating. I. The biometric relations between parents and offspring." *Genetics*, Vol. 6, 1921.

For our problem, these equations may be written:

1.	$d_{SES.MA}$	= c^2	=	.1617
2.	$d_{SES.Y}$	= b^2	=	.8383
				1.0000
3.	$d_{SA.SES}$	= e^2	=	.0081
4.	$d_{SA.MA}$	= h^2	=	.0734
5.	$d_{SA.\overline{SESMA}}$	= $2ceh$	=	.0196
6.	$d_{SA.Z}$	= i^2	=	.8989
				1.0000
7.	$d_{EA.SES}$	= a^2	=	.0056
8.	$d_{EA.MA}$	= d^2	=	.4497
9.	$d_{EA.SA}$	= g^2	=	.0464
10.	$d_{EA.\overline{SASES}}$	= $2age + 2agch$	=	.0064
11.	$d_{EA.\overline{SAMA}}$	= $2dgh + 2dgce$	=	.0887
12.	$d_{EA.\overline{SESMA}}$	= $2adc$	=	.0405
13.	$d_{EA.X}$	= f^2	=	.3627
				1.000

I obtained a check on the accuracy of the computations involved in the second group of equations above by letting i^2 equal $1\text{-}R^2_{SA(MA,\ SES)}$. The value of i^2 is given in Equation 6 as .8989, and the check value is .8990. Similarly, a check on the accuracy of the computations involved in the third group of equations was obtained by letting f^2 equal $1\text{-}R^2_{EA(MA,\ SA,\ SES)}$. Equation 13 gives the value of f^2, which is .3627. The check value is .3624, if R is taken as .7985, and .3629 if it is taken as .7982. These different values for R were obtained by means of two methods of computing R.

Although, in this investigation, we are not concerned with the relative importance of the factors contributing to variability in S.E.S. and S.A., it may, nevertheless, be pointed out that only 16 percent of the variation in S.E.S. is determined by the part of heredity measured by the Binet Scale, while 84 percent is determined by other factors. Similarly, not more than 7 percent of the variation in school attendance is determined by M.A., less than 1 percent by S.E.S. and about 2 percent by S.E.S. and M.A. combined; the remainder, 90 percent, is determined by other factors.

We are chiefly concerned with the third group of equations which gives the percent of variation in E.A. determined by each of the independent variables separately and in different combina-

tions. Only about one-half of 1 percent of the variation in E.A. is determined by S.E.S. directly. Less than 5 percent of the variation in E.A. is determined by S.A., while 45 percent of it is determined by M.A., and 36 percent by other factors. These percentages give the direct influence of the different variables in determining variation in E.A. These variables also have an indirect or joint influence upon variation in E.A. which cannot be separated. The joint influence of S.A. and S.E.S. is only a little more than one-half of one percent, while that of M.A. and S.A. is about 9 percent, and of M.A. and S.E.S. about 4 percent. If the indirect, or joint, influence of M.A. upon variation in E.A. could be separated, it is not unlikely that it would account for at least 50 percent of the variation in E.A. Moreover, what the other influences representing 36 percent are, we do not know. They may be evenly divided among the unaccounted for or unmeasured environmental and hereditary factors. If this assumption is made, then at least two-thirds of the variation in E.A. is determined by the hereditary factor.

If the relative weight of our causal factors is based only upon their direct influence in contributing to variation in E.A., then M.A. has 10 times as much influence as S.A., 80 times as much as S.E.S. and 9 times as much as S.A. and S.E.S. combined. Differences in heredity, then, are 9 times as potent in producing differences in education as are our measured differences in environment. By differences in heredity in this sentence, is meant such as are measured by the Stanford Revision of the Binet Scale; by differences in education, such as are measured by the Stanford Achievement Test; and by differences in the environment, such differences in socio-economic status as are measured by a revised Chapman-Sims Scale, and such differences in training as are represented by differences in school attendance. Moreover, by the word "differences" is meant only such as obtained in the schools and communities involved in the investigation.

It should also be pointed out that these results are no index of the efficiency of the schools and of the home in educating children. They do not show how far the children have been lifted above a zero educational level. They do purport to show under existing educational and socio-economic conditions which of the causal factors is most potent in explaining the enormous differences which are so prevalent in educational achievement.

For the data on E.A., M.A., S.A., and S.E.S. obtained for the ninth-grade girls, Whitney and Woody employed the method of path coefficients. They treated these results in exactly the same way as the results of this investigation were treated, with the exception of correcting for attenuation. I have tabulated here their coefficients of determination, and for the sake of comparison have repeated the coefficients of this study.

	Coefficients of Determination *Whitney-Woody*	*Heilman*
1. $d_{SES.MA}$	.0013	.1617
2. $d_{SES.Y}$	.9987	.8386
	1.0000	1.0000
3. $d_{SA.SES}$	.0023	.0081
4. $d_{SA.MA}$	.0022	.0734
5. $d_{SA.\overline{SESMA}}$	—.0001	.0196
6. $d_{SA.Z}$	.9956	.8989
	1.0000	1.0000
7. $d_{EA.SES}$	.0014	.0056
8. $d_{EA.MA}$	.2532	.4497
9. $d_{EA.SA}$	.0154	.0464
10. $d_{EA.\overline{SASES}}$	—.0004	.0064
11. $d_{EA.\overline{SAMA}}$	—.0056	.0887
12. $d_{EA.\overline{SESMA}}$	—.0014	.0405
13. $d_{EA.X}$	.7374	.3627
	1.0000	1.0000

The most striking difference between these two columns of coefficients is that, for the known variables, those in the first are very much lower than those in the second column. This disparity is in part due to differences in the amount of variation of corresponding variables of the two investigations, and in part to the fact that the Whitney-Woody coefficients are based upon raw correlations, whereas my coefficients are based upon corrected correlations. For this investigation the coefficient of variability for S.A. is 1.6 times the S.A. coefficient for the Whitney-Woody study. The coefficients for the other variables show negligible differences, amounting to about one-tenth of the size of the higher of each pair of coefficients. It is, moreover, significant that in the Whitney-Woody study attendance was reported only in full years; smaller differences in attendance were not recognized. It appears, therefore, that a child

had to attend school only for a month or two in any year to be credited with one year's attendance.

I applied Thomson and Pintner's criterion for the unselected nature of a group to Whitney and Woody's data. The coefficient of correlation for life age and mental age was found to be —.160 and the ratio of the coefficients of variation of life age and mental age .461. Corresponding figures for this investigation are .120 and .142. The requirement is approximate equality of the coefficient and ratio. With these and other rather marked differences between the two investigations, any close correspondence between the two sets of path coefficients should not be expected.

In one respect, however, the two investigations do show a strong tendency in the same direction. Both of them show that mental age has a marked preponderance over all of the other causal factors which appear in the investigations. In fact, this preponderance is relatively much greater for the Whitney-Woody investigation than for this investigation.

Our schools and higher institutions of learning could be very much improved if they were more concerned with the formation of classes upon the basis of equal native endowment than they are with making classifications on the basis of equal training. Curricula are made and administered without any, or very little, reference to individual differences in native talents. I am of the opinion that under existing conditions there is the greatest need for an efficient personnel department in our schools and colleges for the purpose of steering the children and students educationally and vocationally. Such, at least, are the conclusions I have arrived at both from the results of this investigation and from my experiences as a teacher and student in the public schools and in higher institutions of learning.

A Summary of the Main Result, with Some of the Practical Implications

Educational achievement was measured by the Stanford Achievement Test, and was expressed in terms of educational age. In reality, therefore, we determined which of the causal factors, the hereditary or the environmental, were the most potent in producing educational age. The hereditary factors are represented by mental age as measured by the Binet Scale; the environmental fac-

tors are represented by school training as measured by school attendance, and by the socio-economic status of the home as measured by a revision of the Chapman-Sims Socio-Economic Scale.

We found that ten-year-old children differed widely in educational achievement as expressed by the amount of variation in educational age. We also found that these children differed widely in their standing in the causal factors, differences which are expressed in terms of variation. With these facts in hand it became possible to determine which of the causal factors made the largest contribution to the differences in educational achievement.

Our results show that probably 50 percent of the variation in educational age was due to such hereditary factors as had been measured; less than 13 percent of the variation, to school training;[14] not over one percent, to the socio-economic status of the home; and about 36 percent, to such hereditary and environmental factors as had not been accounted for. In schools and communities which differ from the schools and communities from which our data were drawn, the relative significance of these figures would have to be modified. But, as by a similar procedure, almost equivalent results were obtained for all of the ten-year-old children of Greeley, there is some evidence for the belief that they would have to be modified only slightly for different school communities of the Northern states.

I am of the opinion that the results obtained have very significant implications for educational practice in the public schools and in the higher institutions of learning. Our results show that under the prevailing home conditions and school organization, intellectual endowment has by far the most powerful influence in determining differences in achievement in the traditional curriculum. It is not unlikely that a similar statement could be made for achievement in general.

In view of the superior potency of intellectual endowment, it is far more important to make classifications in the public schools and in higher institutions of learning upon the basis of intelli-

---

[14] Thirteen percent represents the sum of the effects of S.A. alone, plus S.A. *in combination* with M.A. It is of interest here to refer to the correlation (corrected for attenuation) between E.A. and S.A., which was .44. The square of this value, or 19 percent, represents the *maximal* contribution which S.A. conceivably makes to variability in E.A., *including* its connection with E.A. through indirect paths of influence, and *regardless* of the philosophical assumptions underlying the use of the path coefficient technique.

gence than upon differences in the amount of training. This untenable custom of classifying upon the basis of the amount of attendance is far more prevalent and more rigidly adhered to in the colleges and universities than in the public schools. It cannot be justified either by the results of investigations or by teaching experiences. Even though the brightest students devote only one-fifth as much time to their studies as the dullest, they still make the higher grades. Thus, for example, bright students who have had no courses in psychology do far better work in that subject than do dull students who have had from two to six courses in the subject. Moreover, in the public schools there are bright children who, in spite of inferior application, complete the work of the elementary school in one-third of the time required by the dull.

A second important inference is that the activities concerned with the educational and vocational guidance of children in the public schools and of students in college should receive more recognition. There should be departments in the school with the functions of classifying children, adapting school tasks and methods of instruction to their learning abilities, and of determining the probable occupational levels at which they could succeed in life.

Efficient personnel departments should also be established in all the higher institutions of learning, charged with the functions of determining what students should be allowed to enter, of making the best classifications under existing practical limitations, and of guiding students into the courses and departments most nearly adapted to their interests and intellectual endowments. Such a department would eventually render more real service to the future citizens of our country than any other department in the institution. As heredity is so much more significant than training in determining differences in educational standing, there should be more effort devoted to guidance, even if less effort must then be devoted to training.

This investigation does not determine how much the children learn in our schools. It does not show how much the schools have contributed in raising the children to their present levels of attainment. But it does offer some evidence in favor of the contention that in human nature there are marked differences in the possibilities of training which must be respected and provided for in every well-regulated school system and higher institution of learning.

PREFATORY NOTE TO CHAPTER III

Miss Denworth, gathering her material in New York, furnishes data which concur very well with those gathered by Heilman in Denver. Although this study and the one by Heilman differ from one another in the age range of subjects, and in the fact that Heilman's study contains data upon the socio-economic status of his subjects while this one does not, both studies agree as to essential conclusions, *i.e.*, in showing a significant relationship between intelligence and school achievement, and in showing only a very slight dependency of deviations in achievement upon variability in school attendance.

Strictly speaking, it should be added that the question whether or not the relationship established between achievement and intelligence directly expresses the causal dependency of school accomplishment upon native capacity is beyond the scope of either the Denworth or the Heilman study to answer. Logically, it is possible that the two factors might have a reciprocal effect upon each other, or both might be affected by some common environmental influence, as far as either study alone could conclude on the basis of its own data.

CHAPTER III

THE EFFECT OF LENGTH OF SCHOOL ATTENDANCE UPON MENTAL AND EDUCATIONAL AGES*

KATHARINE M. DENWORTH
Principal, Bradford Academy, Bradford, Massachusetts

I. THE PURPOSE AND THE NATURE OF THE INVESTIGATION

1. The Importance of the Problem

Ever since schools have existed, it has been tacitly assumed that if only the school and the child could be brought together and kept together long enough, education would result.

Recently, largely as a result of the great individual differences in the mental and the educational status of children of the same age revealed by testing intelligence and achievement, doubt has arisen as to the effectiveness of school instruction in determining mental capacity and educational acquirement. This doubt, persistent and growing, has given rise to numerous attempts, of which the present study is one, to determine to what extent mental and educational development are determined by the efforts of the school and to what extent by mental endowment. Such information is needed to eliminate futile effort and to insure maximal returns on the pupil's, the teacher's, and the public's investment in education.

The best measures of the efficacy of school instruction are the mental and the educational growth which accompany time spent in school. If attendance at school is highly productive of such growth, present school practice may be confidently continued; if unproductive of at least considerable educational development, present practice should be superseded by productive procedure.

---

* This study, undertaken while the author was at Columbia University, has benefited by the encouragement and helpful criticism of Professors Percival M. Symonds, William A. McCall, Rudolf Pintner, William C. Bagley, Thomas H. Briggs, and Bird T. Baldwin (formerly of Swarthmore College).

It owes much to the long-continued coöperation of Dr. Abby Porter Leland, Principal, and Miss Mildred Innes, Acting Vice-Principal, of Public School 157, Manhattan. Coöperation was also extended in carrying out the testing of intelligence by Professor Pintner, his assistants, Miss Carrie Coy and Dr. Gertrude Hildreth, and his graduate students. The scoring of the achievement tests was accomplished through the coöperation of Miss Della Wood.

This investigation, therefore, attempts to determine the effect of time spent in school upon mental capacity and educational acquirement and the influence of mental capacity upon educational acquirement. These effects, however, must be studied in relation to other mental and educational influences.

2. The Delimitation of the Investigation

Of the possible determinants of mentality, this problem deals directly only with attendance at school and age; with the potency of native intelligence it deals only by inference. Of the possible contributors to educational development, the study deals only with intelligence, attendance at school, and age. These factors were selected for study because they are not only important, but are also objective and measurable, and thus lend themselves to scientific investigation.

In as much as group intelligence tests measure many of the same abilities as achievement tests, overlapping of tests was avoided by using as the measure of intelligence the complete form of the Stanford Revision of the Binet-Simon Intelligence Scale—the most valid, reliable, and carefully standardized intelligence test.

The Stanford Achievement Test, a battery of tests covering all the curriculum from the second to the eighth grade, inclusive, was selected as the best measure of all-round educational development. The derivation of these tests and the procedure in administering and scoring them are described in the *Manual of Directions*.[1] This test gives a valid and reliable measure of the educational attainment, not only of the group as a whole, but also of each pupil. The reliability coefficient of the entire battery of tests for unselected pupils of constant age is .98. The probable error of the educational ages, based on the whole test for unselected pupils of ages eight to fourteen, is very low, ranging from 1.5 to 2.1 months.[2]

The total number of days that a child had attended school from his initial enrollment was taken as the most reliable measure of the time he had spent in school.

---

[1] *Stanford Achievement Test, Manual of Directions*. New York: World Book Company, 1924.

[2] *Ibid.*, p. 16.

Within the limits prescribed by the foregoing selection of factors for study and their measures, this investigation endeavors to find tentative answers to the following questions:

To what extent is variation in scores on the Stanford Revision of the Binet-Simon Intelligence Scale determined by variation in length of school attendance? By variation in chronological age?

What influence is exerted upon differences in Stanford Achievement Test scores by differences in length of school attendance? By differences in mental age? By differences in chronological age?

What is the prognostic value of a combination of the Stanford Revision of the Binet-Simon Intelligence Scale, length of school attendance, and age in predicting school success as measured by the Stanford Achievement Test?

3. The Method of Investigation

The ideal way to determine the influence of attendance at school upon intelligence and educational achievement is to use the equivalent-groups experimental method.

From a large number of children who had never had a day of formal instruction at school or at home, two groups would be equated on the bases of chronological age and of intelligence. Then the experimental group would be sent to school regularly; the control group would be given no formal instruction. At the end of a year or more both groups would be given identical tests of intelligence and of educational achievement. Differences between the mental and the educational scores of the experimental and of the control groups could then unquestionably be attributed to the effect of school attendance.

However desirable theoretically, under conditions typical of modern American life such an experiment is obviously impossible. The influences of nature and nurture, which under the relatively constant conditions of public school instruction nevertheless produce radical mental and educational differences in children of the same age, must be studied within more narrow limits. The more accurate equivalent-groups experiment must yield to the more practicable causal investigation.

This causal investigation, consequently, does not treat attendance as either present or absent, but treats of varying lengths of

attendance and of the accompanying variations in mental and in educational development.

Owing to the difficulty of selecting two groups exactly equivalent in educational and chronological ages and two other groups exactly equivalent in mental and chronological ages, the statistical techniques of correlation, partial correlation, and regression equations were employed as satisfactory substitutes.

II. The Sources of the Data and the Procedure in Collecting Them

1. The Subjects

The investigation was made during the spring terms of 1925 and 1926 in Public School 157, Manhattan, New York City, an elementary school with an enrollment of approximately two thousand pupils.

To avoid the influence of accidental selection, since all eight-year-olds have not yet entered school and many thirteen-year-olds have dropped out, only pupils of nine, ten, eleven, and twelve years of age were used in the final results. All the children of these ages, in whatever grade enrolled, were tested only after further selection on two bases. To insure reliable attendance data, pupils were eliminated who had not had all their schooling in the public schools of New York City. To obviate the influence of a language handicap on test scores, children from non-English-speaking homes were also excluded. In the final handling of results pupils were omitted for whom any of the data were missing on account of temporary absence or transfer to other schools.

In the spring of 1925 complete data were secured for 569 pupils. To guarantee greater reliability of the results by increasing the number of cases, data were obtained the following spring for all the new pupils in the school who qualified on the aforesaid bases of age, attendance, and language, and for former pupils who had been too young to qualify the previous year. Complete records were thus procured for 260 additional pupils, making a total of 829 cases.

Included in this number are 52 average and brighter eight-year-olds and 77 average and duller thirteen-year-olds, whose inclusion, however, does not materially alter results, as the influences of the

brighter younger and of the duller older children presumably tend to neutralize each other. Nevertheless, to insure the greatest possible reliability, the final conclusions are based upon only the net total of 700 comparatively unselected subjects, that is, the pupils of ages nine to twelve, inclusive.

The children selected for study have had similar incentives and opportunities to learn. Their social and economic status is rather below the average. Moreover, they have attended only public elementary schools of New York City. They are, therefore, essentially homogeneous with regard to the quality of their school instruction. Their attendance records, however, prove that they are heterogeneous with regard to the quantity of school training. These conditions increase the validity of the study.

2. The Materials and the Procedure

a. The Measurement of Intelligence.—In March, April, and May of both 1925 and 1926 the complete form of the Stanford Revision of the Binet-Simon Intelligence Scale was applied by graduate students of Columbia University who had been trained in intelligence testing by Professor Rudolf Pintner and who had had some experience in testing.

In order not to invalidate results by spurious selection, forty twelve-year-olds who were not tested in 1925 owing to lack of time were tested by the same experienced examiners the following year through the courtesy of Mr. Viertel and Miss Kornman, the principals, respectively, of Public Schools 43 and 136, Manhattan, to which the pupils had been transferred. Early in June, 1926, after the graduate students had discontinued their work, the remaining pupils necessary to complete the age-groups for the second year were tested by Miss Dorothy Wright, a skilled examiner, who generously gave her time.

To render mental and educational data comparable, each child's mental age was computed forward or backward to the date of his achievement test.

b. The Measurement of Educational Acquirement.—The Stanford Achievement Test, Advanced Examination, Form A, was used above the third grade and the Primary Examination, Form A, in the second and third grades. The tests were given within two days

of either April 8 or April 30 in 1925 and within two days of April 22 in 1926.

To secure uniformity in administering and scoring the tests and thus to increase the reliability of the results, the writer, who had had considerable experience in giving and scoring standardized tests, administered all the tests and scored 75 papers. All the other papers were scored and re-checked by an assistant skilled in scoring. Accuracy and consistency were thus assured.

To render the measures of educational attainment comparable with mental and chronological ages, the original total scores yielded by the Stanford Achievement Test were converted into terms of educational age by means of tables in the *Manual of Directions* for the test.

c. *The Measurement of Length of Attendance.*—Since the public schools of New York City have for years kept individual cumulative attendance records, it was possible to procure reliable attendance data.

Only cards were used which unquestionably gave a complete record of a pupil's attendance from his first enrollment in school to the beginning of the semester in which the tests were given. The home-room teacher's daily attendance record supplied the number of days attended from that date to the day of the achievement test. All these data were compiled by the writer with careful checking of each step.

The investigation takes account only of the actual number of days of attendance, disregarding the relative potency of attendance at different chronological ages. For a given pupil one day's attendance at six years of age, for example, will presumably not yield the same mental and educational returns as one day's attendance at twelve years. Obviously, means are needed for equating separately the mental and the educational effectiveness of days of attendance at each chronological age.

To equate days of attendance scientifically for mental and for educational growth at different chronological ages, it would be necessary to give the Stanford-Binet and the Stanford Achievement tests to many children of each age from five to fourteen years. Then, by compiling attendance data for each age-group and checking them against the data of the tests, it would be possible eventually to determine the best mental and educational weights for attendance at each age. The labor involved, however, in the preparatory

steps of giving the Stanford-Binet and the Stanford Achievement tests to many children too young to be used subsequently in the study would render the investigation too onerous to be undertaken.

Since it is better to obtain even a rough measure of the effect of attendance upon mental and educational development than to permit too meticulous a regard for perfection to delay information indefinitely, the use of the actual number of days of attendance, unequated for age differences, seems to be justified as the best empirical measure available. It is believed, moreover, that the handling of attendance data separately for each age-group, the tendency of attendance laws to promote school entrance at nearly the same age for all children, and the effect of chance in balancing inequalities in large groups—it is believed that these considerations eliminate gross inaccuracies in dealing with unequated days of attendance.

d. The Measurement of Chronological Age.—In every case the child's chronological age was reckoned from the date of birth (recorded on his cumulative attendance card) to the date of his educational test.

The subjects having a chronological age from 9 years, 0 days, to 9 years, 11 months, and 31 days, were included in the nine-year-old group, and similarly for the other ages.

III. The Distribution of the Measures

The measurements of all the age-groups, of the 700 unselected cases, and of the entire group of 829 subjects, including the selected eight- and thirteen-year-olds, are summarized in Table I. This table gives the number of cases involved and the means (M), the standard errors of the means (SD of M), the ranges, the standard deviations of the distributions (SD), and the coefficients of variation (V), of all the distributions.

The eight- and thirteen-year-old groups are included in the tables, but not in the discussion of the results.

Table I.—Summary of Measurements

	M	SD of M	Range	SD	V
Age 8 (52 Cases)					
Chronological Age (Months)	105.00		12	2.47	
Mental Age (Months)	121.67		98	20.62	
Educational Age (Months) ..	127.54		60	12.40	
Attendance (Days)	505.80		499	118.34	

TABLE I—*Continued*

SUMMARY OF MEASUREMENTS

	M	SD of M	Range	SD	V
Age 9 (186 Cases)					
Chronological Age (Months)	114.19	.26	12	3.53	3.09
Mental Age (Months)......	118.45	1.15	92	15.64	13.20
Educational Age (Months)..	126.62	1.07	76	14.55	11.49
Attendance (Days)	640.90	10.17	744.5	138.63	21.63
Age 10 (222 Cases)					
Chronological Age (Months)	126.72	.24	12	3.59	2.83
Mental Age (Months)......	127.75	1.54	127	22.96	17.97
Educational Age (Months)..	136.40	1.30	105	19.35	14.19
Attendance (Days)	779.70	10.53	800	156.88	20.12
Age 11 (166 Cases)					
Chronological Age (Months)	138.63	.28	12	3.67	2.65
Mental Age (Months)......	139.12	1.69	124	21.72	15.61
Educational Age (Months)..	150.26	1.49	99	19.26	12.82
Attendance (Days)........	964.50	12.21	911	157.26	16.30
Age 12 (122 Cases)					
Chronological Age (Months)	150.57	.32	12	3.55	2.36
Mental Age (Months)......	153.00	2.69	133	30.17	19.72
Educational Age (Months)..	165.21	1.89	98	21.22	12.84
Attendance (Days)	1146.00	13.35	756	149.82	13.07
Age 13 (77 Cases)					
Chronological Age (Months)	162.26		12	3.49	
Mental Age (Months)......	140.79		102	20.80	
Educational Age (Months)..	155.58		86	17.05	
Attendance (Days)	1253.90		1094	197.68	
Ages 8, 9, 10, 11, 12, 13 (829 Cases)					
Chronological Age (Months)	131.86	.58	72	16.82	12.76
Mental Age (Months)......	132.61	.87	146	25.15	18.97
Educational Age (Months)..	142.59	.79	126	22.67	15.90
Attendance (Days)	868.09	9.39	1370.5	270.39	31.15
Ages 9, 10, 11, 12 (700 Cases)					
Chronological Age (Months)	130.51	.50	48	13.27	10.17
Mental Age (Months)......	132.52	.97	146	25.56	19.29
Educational Age (Months)..	142.27	.87	126	23.02	16.18
Attendance (Days)	852.60	8.85	1240.5	234.26	27.48

1. The Central Tendency of the Measures

The slight excess of the mental over the chronological mean, revealed by Table I in every group, favors the initial assumption that the subjects of this study are, collectively, of average mentality.

The striking superiority of the educational to the mental means may be thought to indicate that for these children acquisition is greater than seeming capacity; that is, to indicate that the school is not only utilizing all the mental capacity of its pupils, but that it is also spurring them on to levels of achievement nearly a year in advance of 'reasonable' expectation. It seems more plausible, however, to attribute the ascendancy of the educational means to an overestimate, by the test-builders, of the educational ages corresponding to total scores on the Stanford Achievement Test.

The low standard errors of the means reported in Table I indicate that, with the possible exception of the mental mean of the twelve-year-olds, much confidence may be placed in the obtained means.

2. The Variability of the Measures

The outstanding feature of Table I is the wide range of mental age within a single chronological age. The brightest twelve-year-old, for example, is 11 years, 1 month, older mentally than the dullest. Less remarkable, but almost equally significant, is the range of educational achievement within any one age. The most advanced ten-year-old, for instance, is 8 years, 9 months, older educationally than the most backward. The attendance scores are, also, highly heterogeneous. On the basis of 200 days to a school year, one twelve-year-old has attended school 4 years, 5 months, 11 days, longer than another.

The size of the standard deviations of mental age shown in Table I substantiates the evidence of the ranges as to the heterogeneity of mentality. The standard deviation of mental age for the twelve-year-olds (30.17 months) is 5.17 months in excess of the standard deviation of mental age for unselected twelve-year-olds. The greater mental variability of these twelve-year-olds accounts for the larger standard error of their mental mean. The standard deviations of educational age, as well as the ranges, denote considerable educational heterogeneity. The large standard deviations

of the attendance distributions warrant the preliminary supposition that the pupils of a single age have had quite different amounts of school instruction.

The inclusion of the eight- and thirteen-year-olds in the total results leaves the variability of the mental and the educational distributions practically unchanged, but considerably increases the variability of attendance—results in conformity with logical deductions. The mental and educational scores of the brighter younger and the duller older children tend to fall within the range of scores of the unselected cases; whereas the attendance data of these younger and older pupils extend both limits of the scale.

3. Comparison of the Groups

As the standard deviation and the range are measures only of absolute variability within a distribution, the variability of the groups was rendered comparable by determining the coefficients of variation, V,[3] of each measured factor for each group.

The variability of attendance, as Table I shows, is highest for the youngest pupils and decreases with each increase of age. Since nine years is probably the earliest age at which all pupils are enrolled in school, length of attendance varies most at this age and less from year to year thereafter.

Mental age, on the contrary, varies least for the nine-year-olds and most for the twelve-year-olds. Evidently, for these twelve-year-old children increments of a comparatively constant common nurture, that is, amounts of attendance upon instruction in the public elementary schools of New York City, have tended not to decrease, but to increase individual variation in intelligence.

Variability in educational attainment, unlike attendance and mentality, shows no decided trend with age. It is, however, lower for the eleven- and twelve-year-olds than for the ten-year-olds. The decrease in the educational variation of the older pupils, which in the twelve-year-old group accompanies an increase in mental variation, and the smaller educational than mental variation at every age suggest that present instructional effort tends to level educational differences among pupils. If this inference is correct, it raises a host of questions—irrelevant to this problem, but highly

---

[3] $V = \frac{100\ SD}{M}$

pertinent to education—as to the desirability of uniformity in educational products.

As the most variable group in attendance is the least variable in mental and educational ages, and the least variable group in attendance is next to the most variable in educational age and the most variable in mental age, it is apparent that differences in the amount of schooling of pupils of a single age do not account for differences in their mental and their educational development.

The nine-year-olds are the most homogeneous of the groups in mental and educational ages, but the most heterogeneous in attendance. The ten-year-olds, the most variable educationally, are on the whole the most variable of the groups; the eleven-year-olds, the least variable. The twelve-year-olds, the most variable in mental age, are the least variable in attendance.

The relative variability of the groups affects the next phase of the investigation, the degree of correspondence between the factors.

IV. The Relationship of the Measures

The relationship of the measures can be adequately considered only in the light of the number of cases, the degree of selection involved, and the variability of the groups studied. As these spurious determinants of correlation have been treated in preceding sections, the study is concerned with the correspondence existing between one series of measures and another. It is interested primarily in the relation of length of school attendance to mental and to educational development, and in the relation of intelligence to educational achievement; and secondarily in the effect of chronological age upon mental and educational ages and upon attendance.

From the frequency distributions for each group scatter diagrams were plotted and converted into correlation tables, from which the coefficients of correlation were determined by Toops's adaptation of Pearson's product-moment formula.[4]

Linearity, required by Pearson's product-moment formula, was assumed throughout. This assumption is partially justified by the prevalence of straight-line relationships in psychological studies

[4] Toops, Herbert A. "Eliminating pitfalls in solving correlation." *Jour. of Expt. Psych.*, 4:434-447.

and by the tendency of the means of the successive arrays of the correlation tables of this study to form straight lines.

The correlations reported are uncorrected for attenuation. Since errors due to chance lower correlation, the coefficients would be higher if all variable errors were eliminated. The formula, however, for correction for attenuation requires not only approximate linearity, but also lack of correlation of the errors with the measures and with each other. These conditions are so seldom realized in practice that corrected coefficients would probably have been too high and would have invalidated the resulting partial correlations.

The use made of partial correlation in causal investigations has been frequently challenged. Miss Burks, in a penetrating criticism of the technique of partial correlation, inveighs against holding constant, variables which may be wholly or partly caused by either of the correlated factors. She, however, concedes the validity of using partial correlation to eliminate the influence of age on the true relationship between two correlated measures:

> "There are situations in which a variable is indisputably a cause rather than an effect. In one obvious type of situation, chronological age is such a variable. The many studies which have employed partial correlation technique to eliminate the contribution of maturity to correlated measures are apparently on safe ground, provided they have published the precise age ranges of their subjects."[5]

Since unquestionably age is not the result, but to some extent the cause of intelligence, educational achievement, and length of school attendance, the use of partial correlation in this investigation to eliminate the influence of variation in age is believed to be entirely valid. *The conclusions of the investigation and their practical applications to education are based upon only the correlations in which age is constant.*

Partial correlations for two variables constant are also reported in the tables because they are needed in the construction of regression equations. Since no attempt is made to interpret these correlations as indicating degrees of causal relationship, the use made of them seems to be defensible.

---

[5] Burks, Barbara S. "On the inadequacy of the partial and multiple correlation technique." *Jour. of Educ. Psych.*, 17:538.

The total correlations, the probable errors of the coefficients, and the partial correlations of the four variables are given for the single age-groups and for the two aggregate groups. Total correlations are shown for the eight- and thirteen-year-olds, but are not discussed.

Since heterogeneity increases correlation, in studying the correlation coefficients reported it should be borne in mind that coefficients, regardless of the existing amount of direct relation, tend to be higher in the two composite age-groups and in the more variable of the single age-groups.

1. The Effect of Length of Attendance upon Mental Age

The interaction of length of attendance and mental age is shown in Table II, which summarizes this relationship for all the groups.

Although attendance and mental age have both varied greatly, the total correlations of these factors in the small groups are negligible. In the youngest and oldest groups the correlation coefficients are, respectively, $.20 \pm .05$ and $.14 \pm .06$. In the eleven-year-old group, correlation rises to $.30 \pm .05$; in the ten-year-old group it reaches $.34 \pm .04$, a reliable but low coefficient. Heterogeneity of age produces total correlations of $.50 \pm .02$ for the restricted and $.46 \pm .02$ for the unrestricted total—substantial correlations with low probable errors.

TABLE II.—TOTAL AND PARTIAL CORRELATIONS BETWEEN SCHOOL ATTENDANCE AND MENTAL AGE.

Age	Number of Cases	Total Correlation	P.E.r	Chronological Age Constant	Chronological and Educational Ages Constant
9	186	.197	.05	.135	.041
10	222	.337	.04	.292	-.079
11	166	.300	.05	.282	.084
12	126	.138	.06	.118	.094
Restricted Total	700	.503	.02	.213	-.030
8	52	.046	.09		
13	77	.107	.08		
Entire Group	829	.481	.02	.199	-.047

When the influence of differences in age is excluded, the correlations are only slightly reduced for the small groups, but are diminished for the composite groups to inconsiderable, although reliable, correlations of $.21 \pm .02$ and $.20 \pm .02$.

Evidently, for the subjects of this study variation in mental development is practically independent of variation in length of school attendance.

2. The Effect of Length of Attendance upon Educational Age

The degree of dependence of educational age upon length of attendance in the various groups is indicated by Table III.

TABLE III.—TOTAL AND PARTIAL CORRELATIONS BETWEEN SCHOOL ATTENDANCE AND EDUCATIONAL AGE.

Age	Number of Cases	Total Correlation	P. E.r	Chronological Age Constant	Chronological and Educational Ages Constant
9	186	.193	.05	.149	.076
10	222	.435	.04	.403	.300
11	166	.335	.05	.313	.164
12	126	.215	.06	.203	.191
Restricted Total	700	.617	.02	.295	.211
8	52	.096	.09		
13	77	.296	.07		
Entire Group	829	.608	.01	.291	.222

In the homogeneous age-groups correlation is present, but not impressive. It varies from $.19 \pm .05$ and $.22 \pm .06$, respectively, in the youngest and oldest groups to $.33 \pm .05$ and $.44 \pm .04$, respectively, in the eleven- and ten-year-old groups. The higher correlation of the ten-year-olds is probably due to greater variability and some selection. A substantial correlation, moreover, of $.62 \pm .02$ for the 700 pupils is confirmed by one of $.61 \pm .01$ for the 829 cases.

Freeing the total correlations from the effect of age variation only slightly decreases correlation for the groups of one age, but halves it for the combined groups; the correlation for the 700 comparatively unselected subjects is $.30 \pm .02$—a reliable, but by no means imposing, correlation.

Apparently, differences in length of school attendance contribute but an inconsiderable portion to differences in educational acquirement.

3. The Effect of Mental Age upon Educational Age

The agreement is marked between mental and educational ages in Table IV, which epitomizes this relationship for all the groups. Pupils with a high mental score tend to have a high educational score.

TABLE IV.—TOTAL AND PARTIAL CORRELATIONS BETWEEN MENTAL AND EDUCATIONAL AGES.

Age	Number of Cases	Total Correlation	P. E.$_r$	Chronological Age Constant	Chronological Age and Attendance Constant
9	186	.723	.02	.717	.711
10	222	.837	.01	.826	.809
11	166	.731	.02	.727	.701
12	126	.834	.02	.833	.832
Restricted Total	700	.841	.01	.781	.770
8	52	.740	.04		
13	77	.621	.05		
Entire Group	829	.825	.01	.780	.770

The total correlations for the groups of one age are remarkably high. That they are reliable is evidenced by the low probable errors of the coefficients. The size of the coefficients varies from $.72 \pm .02$ and $.73 \pm .02$ for the nine- and eleven-year-olds to $.83 \pm .02$ and $.84 \pm .01$ for the twelve- and ten-year-olds. The fact that these latter correlations for groups of homogeneous age are as high as the coefficients, $.84 \pm .01$ and $.82 \pm .01$, for the composite groups demonstrates that the impressive correlations are little affected by differences in age.

The persistence of high correlations when the influence of variability in age is controlled offers confirmatory evidence of this fact. For the groups of single age, eliminating age-variation leaves correlations ranging from $.72 \pm .02$ to $.83 \pm .02$. For each aggregate group, rendering age uniform leaves a surplus score of correlation of $.78 \pm .01$, a very significant correlation.

The high correlation coefficients of mental and educational ages, uninduced by age differences, indicate that variation in mental endowment very materially affects variation in educational achievement. They signify, also, that the Stanford Binet is a good measure of the phases of intelligence necessary for scholastic success.

4. The Effect of Chronological Age upon Mental and Educational Ages and Length of Attendance

The portion that chronological age contributes to mental and educational ages and to length of attendance for each group is shown in Table V by the total correlations of chronological age paired in succession with each of the other variables.

TABLE V.—TOTAL CORRELATIONS BETWEEN CHRONOLOGICAL AGE AND MENTAL AGE, EDUCATIONAL AGE, AND SCHOOL ATTENDANCE.

Age	Number of Cases	Mental Age		Educational Age		School Attendance	
		Total Correlation	P.E.r	Total Correlation	P.E.r	Total Correlation	P.E.r
9	186	.197	.05	.151	.05	.377	.04
10	222	.282	.04	.241	.04	.228	.04
11	166	.117	.05	.141	.05	.235	.05
12	126	.160	.06	.109	.06	.147	.06
Restricted Total	700	.498	.02	.603	.02	.777	.01
8	52	—.119	.09	.076	.09	—.152	.09
13	77	.015	.08	.009	.08	.337	.07
Entire Group	829	.435	.02	.572	.02	.830	.01

The degree of correspondence of chronological with mental age revealed in Table V is positive, but slight, in the single age-groups and is higher for the younger than for the older groups—ranging from $.12 \pm .05$ for the eleven-year-olds to $.28 \pm .04$ for the ten-year-olds. The relationship between chronological age and mental age is probably most naturally represented by the total correlations, which are $.50 \pm .02$ for the total of 700 and $.43 \pm .02$ for the entire group in which selection plays some part. This marked and reliable correlation of .50 for the relatively unselected subjects is in accordance with the facts known about the effect of age upon mental development.

The correspondence of chronological age with educational age shown in Table V is low, with high probable errors for the homogeneous age-groups; it ranges from $.11 \pm .06$ to $.24 \pm .04$. For the heterogeneous age-groups it is $.60 \pm .02$ and $.57 \pm .02$—both very considerable and reliable correlations.

The total correlations between chronological age and attendance given in Table V decrease from $.38 \pm .04$ for the youngest pupils to $.15 \pm .06$ for the oldest. In the heterogeneous groups correlation rises to $.78 \pm .01$ for the restricted, and to $.83 \pm .01$ for the

unrestricted total. These latter correlations for pupils of various ages conform to the well-known fact that during the period of compulsory attendance the length of time a child has attended school depends largely upon his age.

In sum, then, comparison of the coefficients of educational age correlated with mental age, with attendance, and with age indicates that intelligence is much more powerful than either length of school attendance or age in determining educational acquirement. The contrast in the size of these coefficients is particularly marked in the small groups, where variability of age, which influences correlation, is largely controlled.

5. A Test of Selection in the Groups

Since the degree of selection of the subjects affects the size of correlation coefficients, it is important to apply an objective test of the selection involved. Now that the coefficients of variation and the coefficients of correlation have been presented, it is possible to make the empirical test of selection suggested by Godfrey Thomson and Rudolf Pintner.[6] In an unselected group the correlation coefficient of chronological and mental ages, they point out, should be approximately equal to the ratio of the coefficients of variation of chronological and mental ages. Table VI compares the relevant coefficients with the ratios for the four age-groups and for their total.

TABLE VI.

COMPARISON OF THE CORRELATION COEFFICIENTS OF CHRONOLOGICAL AND MENTAL AGES WITH THE RATIO OF THEIR COEFFICIENTS OF VARIATION

Age	Coefficients of Total Correlation	Ratio of Coefficients of Variation
9	.20	.23
10	.28	.16
11	.12	.17
12	.16	.12
Restricted Total (700 cases)	.50	.53

The correspondence of the coefficients of correlation with the ratio of the coefficients of variation, for chronological and mental

[6] "Spurious correlation and relationship between tests." *Jour. of Educ. Psych.*, 15:442.

ages, is so close for the composite group and for the single age-groups—except the ten-year-old group where the difference is .12—that, judged by this relationship, the 700 subjects are, as originally assumed, practically unselected.

The fact that, according to this criterion, the ten-year-olds are the most, and the nine-year-olds the least selected, accounts in part for the respective higher and lower correlation coefficients of the two groups. In the former group more selection, reinforced by higher variability, increases the obtained correlation; in the latter group less selection, supported by lower variability, depresses the obtained correlation.

Judged by this test, selection has played little part in determining the degree of relationship between the variables studied.

V. The Value of the Factors in the Prediction of Educational Age

1. The Regression Equations

Partial correlation is valuable not only in freeing correlation of extraneous influences, which, by affecting the homogeneity of the subjects, distort the real relationship, but also in building regression equations upon which prediction may be based.

The regression equation used in this investigation admits of prophecy as to a pupil's educational achievement from knowledge of his intelligence rating, the length of time he has attended school, and his age. It combines these data so as to yield the most accurate educational score possible from a combination of the other data.

The regression equations[7] for the prediction of *standard* educational scores for the ten-year-old group and for the restricted total of 700 pupils, are, respectively,

$$\bar{z}_1 = .783\ z_2 + .176\ z_3 - .021\ z_4, \text{ and}$$
$$\bar{z}_1 = .692\ z_2 + .171\ z_3 + .126\ z_4.$$

[7] Derived from the equation,

$$\bar{z}_1 = \beta_{12.34} z_2 + \beta_{13.24} z_3 + \beta_{14.23} z_4$$
$$\left(\beta_{12.34} = \frac{\sigma_2}{\sigma_1} b_{12.34} \right)$$

$\bar{z}_1$ designates a pupil's standard educational score $\left(\frac{\text{gross score} - \text{mean}}{\sigma}\right)$ and z_2, z_3, and z_4, respectively, designate his intelligence, attendance, and chronological scores, all expressed as standard scores.

The coefficients of z_2, z_3, and z_4 (the regression weights) indicate the weights to be assigned to the intelligence, the attendance, and the age scores, respectively, in estimating a score in educational achievement from these three scores combined.

The second equation gives the best possible forecast, based upon the 700 comparatively unselected subjects, of a pupil's educational age when his mental age, the length of his school attendance, and his age are known.

A comparison of the regression weights of this equation, .692 for mental age, .171 for attendance, and .126 for age, shows that in the prediction of educational age the relative weights of the Binet score, of length of attendance, and of chronological age are 5.49, 1.36, and 1.00, respectively. These values, however, must not be interpreted to mean that, independent of the other variables, mental age is four times as important as attendance and five and one-half times as important as age in determining educational achievement. The regression weights apply only to the relative weights of the other three scores in predicting an educational score; they do not apply to the relative independent contributions of intelligence, attendance, and age to school success.

2. The Reliability of the Prediction

The statistical reliability of a regression equation is determined by the probable error of estimate and by the multiple coefficient of correlation.

The probable errors of estimate of the regression equations of this study[8] are 6.8 months and 7.6 months for the ten-year group and for the 700 unselected subjects, respectively. The chances are even that the error made in using a predicted instead of an actual educational score would be less than 6.8 months for a ten-year-old and less than 7.6 months for one of the group of 700.

The coefficients of multiple correlation[9] for the ten-year-olds and for the restricted total are, respectively, .85 and .87. This means that if the educational ages of the 700 unselected pupils of this study are estimated from the regression equation, the correla-

---

[8] Computed by the formula,
$PE_{(\text{est. } x_1)} = .6745 \times \sigma_{1.234}$

[9] Computed by the formula,

$$R_{1(234)} = \sqrt{1 - \frac{\sigma^2_{1.234}}{\sigma^2_1}}$$

tion between the estimated scores and the scores which these pupils actually obtained when tested with the Stanford Achievement Test will be .87.

The probable errors of estimate and the multiple coefficients of correlation of the foregoing regression equations indicate that the latter are probably more accurate instruments for estimating the educational scores of the pupils studied than the combined judgment of the best teachers, but that they fall short of the accuracy necessary for extensive use in educational prognosis.

3. The Predictive Indexes

The value of the coefficients of correlation in prediction may be estimated from the coefficients themselves without the intervention of a regression equation. The predictive index $(1-k)$, which gives the positive aspect of the coefficient of alienation,[10] represented by k $(k = \sqrt{1-r^2})$, may be utilized for this purpose. It shows the degree of correspondence between an obtained score in one of two correlated traits and a score in the same trait estimated from knowledge of a score in the other trait.

For the group of 700 pupils the predictive indexes based upon the total correlations of intelligence, of attendance, and of age, with educational age are, respectively, .46, .22, and .20. The predictive indexes of intelligence and of attendance with educational age, for uniform chronological age, are, respectively, .37 and .05; and the predictive index of chronological age with educational age, for uniform attendance, is .03.

It is evident that *individual* educational scores cannot be estimated with accuracy from intelligence, attendance, or age data. It is equally obvious that intelligence is a better index of achievement than attendance and age combined.

VI. The Conclusions and Their Implications for Education

1. Summary of the Conclusions

The interpretations which accompany the data presented in the preceding pages are summarized at this point to serve as a foundation for suggestions concerning the educational significance of the results.

[10] Kelley, T. L. *Statistical Method*, p. 173.

The conclusions can be generalized only in the light of the preliminary definitions and the admitted limitations of the study. Within these limits the data reported seem to warrant the following deductions for at least the 700 comparatively unselected pupils studied:

1. Among the subjects of the same age there are great individual differences in mental and in educational ages and in length of school attendance. These differences very probably represent extreme variation in both the nature and the nurture of pupils of uniform age.

2. Of the homogeneous groups studied, the most variable group in length of school attendance is the least variable in mental and in educational ages; the least variable group in attendance is the most variable group mentally and next to the most variable educationally. Apparently, differences in the amount of schooling of these pupils of single age do not account for differences in their mental and their educational development.

3. Although the quantity of school instruction has varied much and the quality little, the negligible correlations between length of school attendance and mental age indicate that the striking individual differences in the mental capacity of these unselected children of constant age are not appreciably due to variation in the amount of school instruction they have received.

4. The low correlations between length of school attendance and educational age signify that the great individual differences in the educational acquirements of these unselected pupils of the same age are not to a significant extent the result of variation in the time they have spent in school.

5. The persistence of high correlations betweens the mental and the educational ages of the children studied, for constant conditions of chronological age, denotes a close and doubtless a causal relationship between their mental capacity and their educational development. The radical differences in the educational achievement of these unselected subjects are probably due to still greater differences in their mental endowment.

6. Variability in age has contributed much less than variability in mentality to the educational status of the pupils tested. The lower correlations of chronological age with mental and with edu-

cational ages for the homogeneous groups show that some of the correspondence for the composite age-groups is due to heterogeneity.

7. The marked relationship between chronological age and length of school attendance indicates that the number of days the subjects have attended school has been largely conditioned by age.

8. Comparison of the results of mental age, of length of school attendance, and of chronological age, correlated with educational age, particularly for the groups of uniform chronological age, shows that intelligence has contributed much more heavily than either attendance or age to the educational achievement of the children studied.

2. The Educational Implications of the Conclusions

If the data for the 700 practically unselected subjects are reliable, and if the conclusions based upon the data are sound, there is a reasonable presumption that both are more universally applicable.

If the correlations obtained are reliable indexes of existing relationships between variation in time spent in school and both mental and educational variation, it is obvious that considerable instructional effort is now wasted and should be redirected along more fruitful lines.

Since differences in mental capacity among children of a single age are practically uninfluenced by varying amounts of school instruction, but are most influential in determining differences in educational acquirement, it is evident that the result of educational effort is largely dependent upon the mental capacity of the pupils. Consequently, the school should take mental capacity into account in instructional planning. Recognition of the uncontrollable, but controlling power of mental capacity in determining educational development will result in readjustment of the whole educational program to individual capacity and needs.

To this end the school should ascertain, by means of intelligence testing, the present mental status of each pupil and upon that should estimate within fairly broad limits his most probable educational possibilities, limitations, and needs. Such diagnosis and prediction are essential for the proper classification of the pupil and for the determination of the best educational treatment to

meet his present and his future needs. It is futile to give instruction without due consideration of a pupil's ability to profit by it and of his probable use for it.

Fuller realization of the existence and consequence of wide mental variation for constant age will result in classification largely on the basis of intelligence and in the differentiation of courses of study and methods of instruction. Such provision for individual mental variation can be made without neglect of the important integrative function of elementary education in a democracy. The necessary integration can be secured through a common core of material for each of the differentiated courses in a subject of study and through classification in the "special" subjects and participation in extra-classroom activities on bases other than intelligence.

Not only should the school take account of the innate differences which condition educational results, but the state, which provides public education, should also take cognizance of them. Whether public education is regarded as an investment by the state to promote its own interests or to develop the highest type of individualized personality, since in either case the product is largely dependent upon mental endowment, public education cannot be better served than by the state's promoting the science of eugenics and disseminating the facts learned in order to improve the mental equipment of school children.

While the results presented indicate that there is little correspondence between variation in school attendance and variation in educational achievement, it should be noted that the correlations reported do not represent the absolute contribution of time spent in school to educational acquisition. In as much as nearly all the pupils studied were classified from the fourth to the seventh grades, inclusive, it is highly probable that some skills—for example, the mechanics of reading—and some facts—the multiplication tables, for instance—having been mastered by practically all the pupils tested, did not considerably affect variation in the scores, although they did contribute greatly to the individual scores. Had some children been tested who had attended school for only a short time or not at all, the lack of the educational elements common to the pupils who had attended for some time would very materially have lowered their scores, thereby increasing variability and raising the correlation.

In considering the data reported, the limitations of the problem must also be kept in mind. Not all the ends of education have been measured, only information and skills, the narrower and more tangible objectives. Accurate means of measuring the acquisition of ideals, attitudes, and appreciations, the wider and more intangible values of education, are not yet available. Hence this study has made no attempt to measure these broader phases of behavior and can, therefore, shed no light on the extent to which the school is achieving these higher aims.

Notwithstanding the foregoing extenuations, this investigation indicates that the school has made only a beginning in utilizing the mental capacity at its disposal. However largely educational acquirement may be prescribed by mental endowment, it is probable that instructional capacity and effort will always be taxed to the utmost to develop all the educational possibilities of all the pupils. Far from discouraging educational activity, therefore, the results of this investigation present a stimulating challenge to renewed and better directed effort.

BIBLIOGRAPHY

BAGLEY, W. C. *Determinism in Education.* Baltimore: Warwick and York, 1925. 194 pp.

BURKS, BARBARA S. "On the inadequacy of the partial and multiple correlation technique, Part I." *Jour. of Educ. Psych.*, 17: Nov., 1926, 532-540.

BURT, C. *Mental and Scholastic Tests.* London: P. S. King and Son, Ltd., 1921. 648 pp.

COBB, MARGARET V. "The limits set to educational achievement by limited intelligence." *Jour. of Educ. Psych.* 13: Nov. and Dec., 1922, 449-464, 546-555.

COURTIS, S. A. *Why Children Succeed.* Detroit: Courtis Standard Tests, 1925. 271 pp.

GORDON, H. *Mental and Scholastic Tests among Retarded Children.* (Educational Pamphlets, No. 44.) London: Board of Education, 1923. 92 pp.

HOLZINGER, K. J., and FREEMAN, F. N. "The interpretation of Burt's regression equation." *Jour. of Educ. Psych.* 16: Dec., 1925, 577-582.

KELLEY, T. L. "Educational determinism." *Jour. of Educ. Research*, 8: June, 1923, 10-19.

KELLEY, T. L. *The Influence of Nurture upon Native Differences.* New York: The Macmillan Company, 1926. 49 pp.

PINTNER, R. *Intelligence Testing, Methods and Results.* New York: Henry Holt and Company, 1923. 406 pp.

TERMAN, L. M. "The possibilities and limitations of training." *Jour. of Educ. Research*, 10: Dec., 1924, 335-343.

TERMAN, L. M. "The psychological determinist; or democracy and the I.Q.," *Jour. of Educ. Research,* 6: June, 1922, 57-62.

TERMAN, L. M., and others. *Genetic Studies of Genius.* Vol. I. Stanford University Press, 1925. 648 pp.

WHIPPLE, G. M. "Educational determinism; a discussion of Professor Bagley's address at Chicago." *School and Society,* 15: June, 1922, 599-602.

WHIPPLE, G. M. "Endowment, maturity, and training as factors in intelligence scores." *Scientific Mo.,* 18: May, 1924, 496-507.

PREFATORY NOTE TO CHAPTER IV

This chapter may be considered by the reader along with the two that precede it. The evidence presented by Obrien indicates that the cumulative effect upon pupils' school abilities of attending school only seven, rather than eight, months a year amounts to very little, even in the course of eight years, although pupils attending eight-months' schools have had, by the time they reach the eighth grade, the equivalent of one school year more of schooling than those attending the seven-months' schools.

Obrien also reports practically no correlation for several hundred pupils between their percentage of attendance over a period of two and a half years and their scores on achievement tests; but this result cannot be clearly interpreted, owing to the fact, pointed out by the author, that a much larger percent of the group having lowest attendance are chronologically retarded. Thus, the pupils with poorest attendance records for the two and a half years, being older on an average than the classmates with whom they are compared, may actually have had more months or years of schooling than those with better attendance records.

The rather striking conclusions of this study should not be interpreted as meaning that it does not matter how long children go to school. To what extent the children in seven-months' schools or the children with irregular attendance may suffer in the acquirement of ideals, ambitions, and personality or character traits not measured by tests, but cultivated in schools, this study can, of course, give no inkling. If this aspect of school life is not kept in mind, the reader may find himself wondering what, after all, the real purpose of our schools should be.

CHAPTER IV

THE CONDITIONAL VALUE OF A LONGER SCHOOL YEAR IN ONE-TEACHER SCHOOLS

F. P. Obrien
Director, School Service and Research,
University of Kansas, Lawrence, Kansas

Nature of the Investigation

From data gathered in connection with a recent study of pupil achievement in one-teacher schools in Kansas a comparison was made of the achievement of 861 pupils attending one-teacher schools for eight months of each year with the achievement of 334 pupils who were attending one-teacher schools in the same region for seven months of each year. An analysis was also made of the effect on school achievement of different degrees of regularity in school attendance in these one-teacher schools. A total of 98 schools in 14 counties located in different parts of the state were included.

The Tests

The tests employed in this study were: the Courtis Arithmetic Test in Fundamentals, Series B; Burgess Silent Reading Test, Form 1; Buckingham-Ayres Spelling Scale, List W, 20 words; Willing Composition Scale; Ayres Handwriting Scale; and the National Intelligence Test, Scale A, Form 1. All tests were given in Grades III to VIII, except spelling, which was given only in Grades VI, VII, and VIII. Uniform testing procedure was followed, and the work was done under close supervision.

Comparison of Seven-Month and Eight-Month Schools

The pupils in the eight-month schools have had, by the end of the seventh grade, seven months more of schooling than the pupils who attended the shorter session. In Grades VII and VIII the pupils attending schools having the longer session showed a marked tendency toward higher achievement scores. Yet in no subject or grade was the difference equivalent to a half-year of progress.

Below the seventh grade the difference in median scores, as shown in Table I, favored the eight-month schools in only nine of the twenty subject and grade comparisons. There was no difference in two of the comparisons.

TABLE I.—COMPARISON OF THE PUPILS' MEDIAN SCORES BY GRADES IN SEVEN-MONTH AND EIGHT-MONTH SCHOOLS OF THE ONE-TEACHER TYPE

Grade	Spelling		Reading		Composition	
	Seven-Month Schools	Eight-Month Schools	Seven-Month Schools	Eight-Month Schools	Seven-Month Schools	Eight-Month Schools
III	....	....	2.3	1.8	16.4	16.3
IV	....	....	3.6	3.7	22.6	20.6
V	....	....	4.7	5.2	25.6	27.2
VI	3.7	3.1	5.2	6.6	29.1	29.0
VII	4.3	6.5	6.4	7.8	31.9	36.9
VIII	10.0	9.5	8.5	8.6	34.0	37.8
Number of pupils	190	439	295	759	334	861
Total	629		1,054		1,195	

Grade	Handwriting		Addition		Multiplication	
III	23.1	23.7	0.7	0.7	....	....
IV	23.2	24.8	1.1	1.5	0.6	0.8
V	26.3	29.0	1.6	1.5	2.1	1.6
VI	31.8	31.1	2.5	2.4	2.8	2.8
VII	35.4	37.1	2.8	3.0	3.8	3.7
VIII	38.6	39.5	3.8	3.7	4.9	5.0
Number of pupils	334	861	334	861	276	727
Total	1,195		1,195		1,003	

RELATION OF ATTENDANCE TO ACHIEVEMENT

The general bearing of the facts presented relative to the length of the school year is strikingly reinforced by an analysis of the relation of attendance to achievement of pupils in these one-teacher schools. The school attendance records were secured over a period of two and one-half years for 585 pupils in 63 one-teacher schools.

In this analysis the pupils were classified into three attendance groups: those who had attended 90 percent or more of the time, those who had been present from 80 to 89 percent of the time, and those attending less than 80 percent of the time. These three attendance groups included 333, 149, and 103 pupils, respectively.

It is significant also that pupils in the highest attendance group generally had also the longer school year.

The test results showed that the median achievement of each of the two lower attendance groups frequently equalled or surpassed the median accomplishment of pupils in the highest attendance group. In no grade was the highest attendance group superior in all the tests. Evidently, regular school attendance alone is not so important an influencing factor in school achievement in rural elementary schools as has ordinarily been assumed. Its value is at least conditioned by other factors.

The correlation between the percentage of attendance of these pupils and their mental test scores was close to zero ($r = 0.05$). As might be expected, however, a much larger percent of the lowest attendance group was retarded chronologically and a smaller percent accelerated than was true in either of the other two groups.

PREFATORY NOTE TO CHAPTER V

Chapters V, VI, and VII all deal with the amount of effect produced upon a pupil's educational achievement by the environmental factors teaching ability or school methods. In the first of these chapters is to be found a type of study that is much needed, but very difficult to make. The problem is concerned with the relative contribution of a pupil's intelligence and of the kind of teaching he receives to the scholastic progress he is able to make, in this case to progress in arithmetic fundamentals and reading comprehension.

The problem is a difficult one to handle because we are not sure just what teaching ability really is, nor just how to estimate it; and because intelligence is a permanent attribute showing a cumulative effect throughout a child's school career, whereas teaching influence changes as the child moves from grade to grade.

The first difficulty—that of estimating teaching ability itself—may possibly obstruct a complete study of the problem at hand until a valid method of evaluating teachers is devised by experimental means. Mr. Taylor limits his investigation, however, to a study of the effect of teaching ability as rated by school administrators. Though admitting that such ratings may include "indications of the general reputation a teacher bears for coöperativeness, educational up-to-dateness, and disciplinary success—all of which may or may not be closely related to the measured achievement of the pupils," he justifies his study by pointing out that at least the ratings secured are those commonly used to designate teaching ability, and that it is important and pertinent to investigate the influence of teaching ability thus rated.

The difficulty due to the prolonged and cumulative influence of intelligence and of the short duration of the influence of the particular teacher in question is partly met by computing a regression equation for predicting achievement in arithmetic and in reading from intelligence, overageness, teaching ability, and initial achievement of the pupils in arithmetic and reading at the beginning of the experimental semester. Since initial achievement is itself a product of the *previous* influences upon learning of pupils, the regression coefficients of intelligence, overageness, and teaching ability must in consequence represent the effects of these variables only during the semester measured. As is evident from Tables II and III, estimated teaching ability has very little effect upon the achievement of pupils over a span of one semester.

Although there may be some question regarding the logical basis upon which Mr. Taylor derived the reliability coefficients which he uses in the corrections for attenuation, it should be noted that the conclusions of the study would not be greatly altered if the corrected coefficients were omitted altogether.

CHAPTER V

THE INFLUENCE OF THE TEACHER ON RELATIVE CLASS STANDING IN ARITHMETIC FUNDAMENTALS AND READING COMPREHENSION

HOWARD TAYLOR
University of Oregon, Eugene, Oregon

There are at least two ways in which the influence of the teacher may conceivably be an important factor in determining the achievement of students: (1) personality may have an inspiring motivating effect; (2) instruction may help the learner use better methods. In this study it has been impossible to separate these two influences. Moreover, it may be that the ratings used to measure the capacity of each teacher to influence the achievement of her pupils are not so much ratings of teaching ability as they are indications of the general reputation a teacher bears for coöperativeness, educational up-to-dateness, and disciplinary success—all of which may or may not be closely related to the measured achievement of the pupils. Nevertheless, such ratings of teachers are commonly used as indications of a hypothetical teaching power, and it seemed pertinent to use them in our effort to analyze the factors underlying pupil achievement in two subjects, arithmetic fundamentals and reading comprehension.

For our study we used the performance of pupils during the first semester of the school year 1923-24 and compared it with the ratings of the teachers who were responsible for the instruction during the same period. The records available pertained to 105 classes in the ten different half grades (4B to 8A inclusive) of nine elementary schools of a medium-sized city. They comprised data on the following items:

1. Final achievement in Arithmetic Fundamentals (Woody-McCall, Form I) and Reading Comprehension (Thorndike-McCall, Form II)
2. Initial standing in Arithmetic Fundamentals (Woody-McCall, Form I) and Reading Comprehension (Thorndike-McCall, Form I)

3. Standing in "Overageness" (Chronological Age)
4. Standing in Intelligence (National Intelligence Tests)
5. Teaching Ability to which each was exposed in arithmetic and reading

In some cases, on account of departmental instruction, certain classes had one teacher for arithmetic and another for reading. The rating appropriate to the teaching ability of the actual instructor has been assigned to each class in each subject.

We need, before presenting the results, to explain the general logic of our investigations. At the beginning of the school year each teacher is faced with a roomful of children. During the semester she is undergoing a practical test of teaching ability. Each child in the room constitutes an "element" in the test. But the elements differ from one another. The initial capacity of each pupil, his intelligence, his age in comparison with his grade position, are all presumably related in some degree to his final achievement in skill and knowledge which his teacher for that year or term is striving to make as large as possible. The statistical treatment used in an analysis of this situation is based on the following reasoning:

1. If any of the factors mentioned are important in determining the final standing of a class, they should tend to be operative in similar fashion throughout the grade range 4B to 8A. The instruction is similar and the effort and ability required of pupils is of essentially the same sort in each grade. All our measurements have been made, in fact, by the same standard tests throughout this grade range.

2. The extent to which these factors, when properly weighted, will predict the final standing of a class with reference to the other classes of the same grade is important evidence of the significance of each factor independent of the others.

3. The average performance of all the pupils in a class whose records in each variable factor are available may be taken as representative of the standing of each class in each variable.

Obviously, it is not important to know which classes are farthest above or below the average of the whole group of 105 different classes in the 10 different half grades. What is desired is to ascertain the tendencies toward concomitant variation of these factors within each half grade about the mean of all such half-grade classes in the city. Thus, it is the deviation of each 4B class from the mean of all the other 4B classes in each variable which concerns us. In the same way, we wish to know

whether or not there is any consistent tendency for the other classes of each half grade to fluctuate about the average of each group of classes in the same half grade for each of the factors upon which data have been obtained. Therefore, the relative standing of each class in each variable has been measured from the mean of the 9 or 15 (or whatever the exact number may be) similar classes in the same half grade. This assumes that if a 6A class is a given distance above the average of the other 6A classes in final ability in arithmetic fundamentals, it is an indication of exactly the same superiority in this kind of achievement as in the case of a 4B or 8A class which is an equal distance above the mean of the other classes in its own half grade.

This assumption that the means of several series of measures of the same mental function are equivalent reference points from which to make measurements is a common statistical procedure and would seem to be entirely justified in this case.[1]

It is necessary to make these measurements of the distance each class is above or below the mean of the other classes of the same half grade, in essentially equivalent units. We have no way of knowing that the 5A class which is 6 test-score units above the mean of the other 5A classes in the city is as superior as the 8B class which is, say, 3 test-score units above the average of its half grade. Three test-score units above the mean of 8B classes *may* be indicative of greater, not less superiority than 6 test-score units above the mean of 5A classes. In such cases it is customary to consider some measure of the variability of the group as an equivalent unit of distance from the mean. Throughout this study the standard deviation of the group concerned has been used as the measure of distance above and below the respective means, which have been regarded as the zero points of each distribution of class averages in each half grade.[1]

Thus, starting with the records of each child in six different variables (initial and final achievement in arithmetic, initial and final achievement in reading, intelligence, and chronological age), the standing of each class has been determined from the records of those children in the class for whom there were complete records. The total number of children whose records constitute the original data of this study is 1968. Each class is represented by an average of the records of the children who composed it. In general, there were about 19 children per class for whom all the data were available. In some cases classes were considerably smaller than this; in others considerably larger. Each class average was then expressed as a deviation from the mean of a distribution of similar class

[1] Kelley, T. L. *Statistical Method*, Section 34, p. 114.

averages, including all the other classes of the same half grade throughout the city school system. Each deviation was divided by the standard deviation of the distribution of class averages to which it belonged. Thus, essentially comparable measures of the various factors were secured for 105 elementary-school classes.

In the same way the differences in teaching ability had to be expressed for each class teacher. The data were gained from two sets of rankings by the principals of each elementary school and two sets of ratings made by the head of the city research department, who was acquainted with the classroom work of all teachers. The principals' first ranking correlated with their second, made over a year later, to the extent of $.649 \pm .05$. The two ratings made by the head of the research department correlated $.883 \pm .02$. The research department ratings (sum of two) correlated with the principals' rankings (sum of two) $.625 \pm .04$. This amount of correspondence indicates that to a considerable extent the same factors were operative in each set of estimates, but that the judgments were nevertheless made independently.

Proceeding according to the reasoning outlined above, the rating of each teacher of each class was compared with the ratings of all the other teachers of similar classes in the same subject and within the same half-grade. The average estimated teaching ability assigned to that grade was considered as a reference point, or base, from which to measure the *relative* degree of teaching ability which each class enjoyed. The standard deviation of the estimated teaching ability assigned to the different classes of each half grade was the unit in which these differences in teaching ability were measured. By this method comparable measures of the teaching ability effective in each class in each subject were computed. These measures were computed separately for the sum of the two ratings made by the head of the research department and for the sum of the two rankings made by the principals.

It is important to know the reliability of these sets of measures used to compare each class and each teacher with the other classes and teachers of the same half grade in the city. The reliability was figured in this way. The records of each child in a class constitute a sample of the initial ability in each subject, the degree of "overageness," the degree of intelligence, and the final achievement of

the class working under a given degree of teaching ability. If the standing of each class be determined from a random half of the pupils in it and this standing be compared with the standing of the class determined from the other half of the pupils in it, the comparison will furnish evidence of the consistency with which these two samples measure the differences studied. Accordingly, the records of each alternate child in each class were grouped together (with approximately equal proportion of the sexes in the groups) and the two groups obtained were treated separately, as has been described. The self-correlation of these measures of class standing in each variable and the prophesied reliability[2] of their sum is shown in Table I. The probable errors of these estimated coefficients have been computed from Shen's formula.[3] N in all cases equals the number of classes compared, *i.e.*, 105.

TABLE I.—SELF-CORRELATIONS OF VARIOUS MEASURES

	Self-Correlation	Estimated Reliability	P. E.
Final Arithmetic	.7470	.855	±.02
Final Reading	.6068	.755	±.03
Initial Arithmetic	.6347	.777	±.03
Initial Reading	.6110	.759	±.03
Chronological age (Overageness)	.6345	.776	±.03
Intelligence	.6039	.753	±.03
Teaching Ability of Arithmetic Teachers	.5123	.678	±.04
Teaching Ability of Reading Teachers	.5309	.694	±.04

The estimated teaching ability of each teacher as compared with that of the other teachers of the same subject in the same half grade was determined separately from the principal's ranking and the research department's rating. It is the correlation of these separate determinations which has been used to estimate the reliability of their sum in each case. The result indicates some shrinkage in reliability when the comparisons are restricted to teachers within the same half grade. The original correlation of these two basic measures of teaching ability was $.625 \pm .04$, when the prophesied reliability of their sum would be $.769 \pm .03$. After undergoing the treatment necessary to this problem, where only differences within each half grade are considered, the same variables show

[2] Kelley, *Statistical Method*, Formula 158, p. 206.
[3] Eugene Shen, *Jour. of Educ. Psych.*, 1924, p. 462.

a prophesied reliability of .678 ± .04, and .694 ±.04, respectively, for the slightly different relative array of teachers of arithmetic as compared with teachers of reading. When it is remembered that only the differences within each rather narrow half-grade range are involved, these reliability coefficients would seem to be fairly satisfactory. Further analysis was made by partial and multiple correlation.

The final standing of each class in each subject relative to the other classes of the same half grade has been considered a criterion. The question is: "What is the significance of each factor in predicting which class will stand highest in final achievement?" There are four variables: initial standing in each subject, chronological age (overageness), intelligence, and teaching ability in each subject; each of these is to be correlated with the criterion.

Results for Arithmetic

Table II shows the intercorrelations of zero order and such partial correlations of the 1st, 2nd, and 3rd order as are essential to the computation of the multiple correlation between the dependent variables and the criterion.[4]

The third-order partials show the residual relationships when the other factors have been held constant so far as is possible by this mathematical technique. The regression coefficients b12.345, b13.245, etc., show the weighting to be attached to each factor in order to give the best possible prediction of the criterion. The theoretical importance of each factor in this prediction is shown somewhat better when we set Sigma 1 equal to Sigma 2 equals Sigma 3 equals Sigma 4 equals Sigma 5, and compute the regression coefficients. (Obviously, the actual units in which each variable was measured are not important.) The result indicates that, so far as the unique factors in each variable are concerned, the intelligent class, the initially capable class, the over-age class, and the class with a teacher who is rated high tend to have some advantage in the struggle for above-average final standing in Arithmetic Fundamentals. Intelligence and high initial capacity would seem to be of about equal importance when net relationships within a

[4] Huffaker, C. L., *Jour. of Applied Psych.*, June, 1923, p. 135.

TABLE II.—PARTIAL AND MULTIPLE CORRELATIONS OF FINAL ARITHMETIC STANDING WITH INITIAL ARITHMETIC, OVERAGENESS, INTELLIGENCE, AND TEACHING ABILITY ASSIGNED TO ARITHMETIC CLASSES

Sigma 1 = 1.894 Final Arith.	Sigma 2 = 1.848 Initial Arith.	Sigma 3 = 1.817 Overageness	Sigma 4 = 1.792 Intelligence	Sigma 5 = 1.775 Teaching Ability
	$r_{12} = +.412$	$r_{13} = -.093$	$r_{14} = +.333$	$r_{15} = +.144$
		$r_{23} = -.241$	$r_{24} = +.426$	$r_{25} = +.130$
			$r_{34} = -.626$	$r_{35} = -.235$
				$r_{45} = +.054$

First-Order Partials	Second-Order Partials
r 12.4 = +.317	r 12.45 = +.307
r 13.2 = +.007	r 13.45 = +.200
r 13.4 = +.157	r 14.23 = +.244
r 14.2 = +.191	r 15.23 = +.104
r 15.2 = +.100	r 23.45 = +.071
r 15.4 = +.134	r 45.23 = −.164
r 23.4 = +.037	
r 25.4 = +.119	Third-Order Partials
r 34.2 = −.597	r 12.453 = +.300
r 35.2 = −.211	r 13.452 = +.188
r 35.4 = −.259	r 14.235 = +.266
r 45.2 = −.002	r 15.234 = +.150

$R_{1(2345)} = .486$

Sigma 1.2345 = 1.6747	Sigma 4.1235 = 1.2302
Sigma 2.1345 = 1.5760	Sigma 5.1234 = 1.6792
Sigma 3.1245 = 1.3344	

b12.345 = .3188	or = .311 if sigmas are set equal
b13.245 = .2359	or = .226 if sigmas are set equal
b14.235 = .3621	or = .343 if sigmas are set equal
b15.234 = .1496	or = .140 if sigmas are set equal

single semester only are considered. Over-ageness is less important than these and a highly-rated teacher is still less important.

Superficially, it might be argued that the final standing of an elementary-school pupil in arithmetic computation is the summation of N segments of achievement (where N equals the number of semesters of school life) in each of which about the same proportion might be expected to vary concomitantly with initial score, overageness, intelligence, and teacher influence. Hence, it might be claimed that intelligence and initial score would always predict the major portion of the total arithmetic achievement. Thus, no matter how important the sum of N such semesters of teacher influence might be, there would be N much larger segments of arithmetic achievement paralleling initial standing and intelligence.

There are two fallacies in this point of view.[5] (1) Initial capacity and probably other variables in this regression equation will not maintain their present relative importance throughout the period of elementary-school life. (2) The regression equation is an instrument of prediction, not a process which coagulates or which centrifuges the constituent independent elements of a conglomerate mass. Making allowance for the first type of inaccuracy, one is forced to admit that if intelligence includes general potentiality, there must be a time of zero initial ability in arithmetic computation. Hence, there must be a time in the early life of the child when initial capacity is no indication of his final capacity. Each semester's increment of ability becomes the initial capacity for the next semester. These data merely show that by the time classes get into the later years of elementary-school life (grades 4-8) what they have achieved to date is an important indication of what they will achieve in the following semesters. In the same way, it might be argued that the importance of relative retardation as a positive indication of final achievement is not constant throughout the elementary-school period. Presumably, the variable called intelligence is fairly constant in its relation to achievement, and it may easily be that teacher influence is approximately constant or perhaps even greater in the earlier, more plastic years of elementary schooling than in the final years which have been considered here. Thus, we cannot safely assume that the relative importance of these factors in final achievement during a single semester will be maintained throughout the other semesters of school life.

Still more important in the correct interpretation of these data is the technical nature of the regression equation. It affords an analysis of the relative predictive importance of various factors with reference to a criterion. The principles which determine the predictive importance of each variable can be simply stated.

(1) The more accurate the measure (reliability), the greater its predictive significance.

(2) The higher the correlation of any measure with the criterion, the greater its predictive significance.

[5] I am indebted to Professor T. L. Kelley for calling my attention to the primary importance of these considerations in this particular problem. He should not, however, be held responsible for such errors as may have crept into my use of them in the following interpretation.

(3) The lower the intercorrelations of any measure with all the other measures, the greater the predictive significance.

Thus, it is a sort of average predictive significance, after due allowance for each of these considerations, which is indicated by the regression coefficient of each variable. We may now attempt to estimate roughly what would happen if we had sufficient information to build up a regression equation to predict final achievement at the end of 16 semesters of elementary-school instruction.

There are no very large differences in the accuracy of these various measures (Table I). Hence, while all our measures predict somewhat better than is apparent from the relationships given—partly due to the unreliability of the criterion and partly to the chance inaccuracy in the measures themselves—none of them would gain notably over the others for this reason. But our measures of teacher influence are somewhat less reliable than the others, hence, teacher influence would gain the most in true relative importance from this standpoint.

As has been shown previously, the correlation of initial ability with the criterion, final standing, would tend toward zero in the first semester and probably tend to increase somewhat with each additional semester. What is called 'initial ability' in these data includes all the previous teacher and other influences which have been at work. Hence, the average correlation of initial ability with final standing for the whole sixteen semesters would probably be somewhat smaller than that found for the one semester among upper-grade classes. Moreover, a part of the correlation with the criterion which would be lost by initial ability would in all probability be gained by the factor, teacher influence. It might be surmised that the same thing is to some extent true of overageness as a factor in final achievement. Presumably, intelligence and teacher influence would at least maintain their respective relationships to the criterion throughout the sixteen semesters.

But the application of the third principle will produce very important changes. The average intercorrelation of initial standing for each of sixteen semesters of school life will be fairly high. So also for overageness and intelligence. Each additional semester will add only slightly then to the predictive significance of these factors as indications of the total elementary-school achievement.

But except for teachers transferred to a higher grade along with their classes and except for the effects of departmental instruction where the same teacher had the class through several years, there is probably no correlation between teacher influence in successive semesters. Each semester of teacher influence is an almost entirely unique indication of final achievement. We could expect, then, very radical changes in the relative significance of these four variables as indications of final standing if we consider the interrelationships of all sixteen semesters of elementary-school life instead of the relationships obtaining during a single semester of instruction. In fact, it is conceivable that the cumulative influence of all the different elementary-school teachers with which each child or class comes into contact would have greater weight in a regression equation for the prediction of total final achievement than any other of the four factors studied. Even so, it is worth noting that this relatively stable characteristic of a child or class which we call 'intelligence' would equal in importance the influence of several teachers at least.

It may be, moreover, that the ratings do not identify fully the truly able teacher of arithmetic, as measured by the changed rank of her pupils in final, as compared with initial, standing. Something causes a very marked change in relative class standing between the beginning and end of a school semester. The multiple correlation $R1(2345) = .486 \pm .05$ indicates that only a very moderate proportion of these factors are involved in the four variables considered. The alienation coefficient $k_1(2345) = .874$ indicates that there are factors which, if known, might be expected to correlate to that extent with final achievement. Thus, in terms of the relationship $r^2 + k^2 = 1.00$, we may say that 76.4 percent of the total determining factors in final class standing in arithmetic remain undiscovered while these four variables contain 23.6 percent.

If allowance is made for the reliability of the criterion[6] the result is $r_1\text{infinity} = \frac{.486}{\sqrt{.855}} = .525$, in which case 27.6 percent of the total determining factors are known and 72.4 percent unknown from the variables under consideration.

---

[6] Kelley, *Statistical Method,* Formula 153a.

Results for Reading

The intercorrelations and essential partials of the four variables with final class standing in reading as a criterion are shown in Table III.

Table III.—Partial and Multiple Correlations of Final Reading with Initial Reading, Overageness, Intelligence, and Teaching Ability Assigned to Reading

Final Reading	Initial Reading	Overageness	Intelligence	Teaching Ability in Reading
Sigma 1 = 1.799	Sigma 2 = 1.801	Sigma 3 = 1.817	Sigma 4 = 1.792	Sigma 5 = 1.817
	$r_{12} = +.831$	$r_{13} = -.525$	$r_{14} = +.714$	$r_{15} = -.050$
		$r_{23} = -.617$	$r_{24} = +.772$	$r_{25} = -.148$
			$r_{34} = -.626$	$r_{35} = +.160$
				$r_{45} = -.219$

First-Order Partials	Second-Order Partials
r 12.4 = +.623	r 12.45 = +.626
r 13.2 = −.028	r 13.45 = −.150
r 13.4 = −.143	r 14.23 = +.205
r 14.2 = +.204	r 15.23 = +.135
r 15.2 = +.132	r 23.45 = −.280
r 15.4 = +.155	r 45.23 = −.147
r 23.4 = −.270	**Third-Order Partial**
r 25.4 = +.034	r 12.453 = +.615
r 34.2 = −.299	r 13.452 = +.034
r 35.2 = +.089	r 14.235 = +.230
r 35.4 = +.030	r 15.234 = +.171
r 45.2 = −.166	

$R_{1(2345)} = .844$

Sigma 1.2345 = .9643	Sigma 4.1235 = 1.0465
Sigma 2.1345 = .8663	Sigma 5.1234 = 1.7443
Sigma 3.1245 = 1.3591	

b12.345 = .6846 or .686 if sigmas are set equal
b13.245 = .0241 or .024 if sigmas are set equal
b14.235 = .2119 or .211 if sigmas are set equal
b15.234 = .0945 or .095 if sigmas are set equal

Although in the zero-order correlations there was a slight negative relationship between final achievement in reading and teaching ability ($r_{15} = -.049$), this was probably due to an apparent tendency to assign initially poor and less intelligent classes to good teachers ($r_{25} = -.148$ and $r_{45} = -.219$). And indeed, the residual correlation of estimated teaching ability with final class standing in reading is positive and significant, though somewhat smaller than the residual relationship of intelligence to the criterion. Turning

to the regression coefficients, it is clear that initial ability comprises by far the most important group of unique factors on which to base a prediction of final class standing in reading; that retardation as measured by chronological age, though positive, is insignificant as a factor. Intelligence has a weight of only one-third that of initial standing; and teaching ability, a weight of about half that given intelligence. The theoretical weights secured by setting Sigma 1 equal to Sigma 2 equals Sigma 3 equals Sigma 4 equals Sigma 5, are practically identical with the others.

Here, again, the interpretation of the facts is subject to the considerations already stated with reference to the previous regression equation. Much of the importance which initial standing has as an indication of the next semester's achievement in reading would disappear in an equation to predict total elementary-school achievement in reading. Indeed, no small portion of it would probably gravitate toward teacher influence. And, on account of the much lower intercorrelations of successive semesters of teacher influence, that variable would tend to accumulate weight more than the others. Hence, here again, teacher influence would probably acquire more weight in a regression equation for the prediction of all sixteen semesters of achievement in reading than would any of the other factors considered. However, it would probably require the influence of more than a single teacher to match, in significance for final reading achievement, the personal characteristics subsumed under the head of intelligence.

In the case of final class standing in Reading Comprehension, the multiple correlation is $R_{1(2345)} = .844$, and $k_{1(2345)} = .536$. Thus, the four factors under consideration give a much better prediction of final standing in Reading Comprehension than of final standing in Arithmetic Fundamentals. Indeed, the proportion of known to unknown factors is almost exactly reversed;[7] 71.2 percent of the total determining factors in final class standing in reading are given by the four variables under consideration, while 28.8 percent remain undiscovered. If the reliability of the criterion is taken into consideration, this particular combination of four factors would predict a true criterion of final class standing in reading to the extent of $r_1$infinity $= \frac{.844}{\sqrt{.755}} = .971$, or 94.3 per-

[7] Utilizing the relationship $r^2 + k^2 = 1.00$.

cent of the total determining factors to 5.7 percent of unknown factors. Thus, it is clear that a true measure of class standing after a semester of instruction in reading could be almost perfectly predicted from a knowledge of initial ability, overageness (C.A.), intelligence, and the teaching ability to which the class was exposed. We can be sure, therefore, that our ratings have not failed to identify most of the teacher qualities which are involved in the relative reading achievement of these classes. Even if all the unknown factors were simply unrecognized teaching talent, the influence of a single upper-grade teacher would remain in general a minor factor in reading achievement. In computation ability this is not so certain. It is quite conceivable that the unknown factors in this sort of arithmetic achievement are in some degree specialized characteristics of good arithmetic teachers which are unnoticed in such ratings as have been used.

Conclusion

In conclusion it might be said that the final standing of an elementary-school class in relation to the other classes of the same half grade in a representative school system varies directly as the initial standing of the class is high or low, as the class is relatively overage for the grade or underage, as it is intelligent or unintelligent, and as it is taught by a teacher who is esteemed by her superiors or by one whose competence is not so well attested. Final achievement is a complex in which many factors are blended. Everything counts, but nothing superlatively. Teachers are not miracle workers. In general, pupils and classes are not transformed in one semester or even in several. But there does seem to be a sort of differential pressure varying with the estimated ability of the teacher which, quite independently of other factors, parallels the variations in final achievement. It effects only very slightly the outcome of each semester of instruction, but its unique character suggests that the sum total of teacher influence on a given child or class would constitute very important data for the prediction of total elementary-school achievement. This uniqueness (approximate zero correlation between successive semesters or years) of teacher influence is probably the reason it is difficult to identify as a factor in achievement under ordinary school condi-

tions of changing teachers. In general, each child or class in the course of an eight-year period is exposed to approximately average teaching ability, since there are about as many poor as good teachers, and the effects of inferior instruction tend to cancel the effects of superior teaching in each individual case. Thus "quality of teaching received" is a fluctuating possession of a child or class, one almost impossible to identify, whereas intelligence and attainment-to-date, being present each year to about the same degree, are increasingly stable possessions which can be readily recognized and evaluated.

Of all the environmental influences by which we hope to improve humanity we are prone to depend most upon the teacher. Is it not significant that, as early as in the upper grades of the elementary school, if we consider only such relative achievement as actually takes place during one semester of instruction in reading comprehension, the influence of the teacher affords only a minor indication of the final standing in comparison with the influence of previous attainment and is far less important than that characteristic of each individual child (intelligence) which has attained such stability at three years of age as to justify predictions of future achievement? Nature and nurture seem to be inextricably intertwined, but he who would change markedly an individual's relative standing in school achievement by such teacher influence as has been measured here must not only begin early, but he must also insure a continuity of superior instruction. The line between potentiality and achievement cannot be clearly drawn. But when early achievement becomes so heavily fraught with the promise of what final standing will be, and such highly esteemed environmental factors as the teacher's influence give an indication of final semester standing which is slight in comparison with that afforded by measures of the mental organization that has already come to pass, are we not forced to think of each new environmental influence as a factor of diminishing import in comparison with the individual as he is at any but the very earliest stages of development?

PREFATORY NOTE TO CHAPTER VI

The problem of the following chapter is the relative importance of intelligence and grade placement as determiners of school achievement. Van Wagenen compares (a) groups of boys of the same mental age, but two years apart in grade placement with (b) boys in the same school grade, but two years apart in mental age. He finds a larger disparity of school achievement in the latter comparison for all the school abilities tested except history information—a subject more emphasized in the higher grades. In reading, the disparity in the second comparison was more than two and one-half times as large as in the first comparison, and in the 'thought phase' of history the ratio of the disparities was nearly as great.

It is hardly possible to state the results of this study in *numerical* terms of the relative importance of intelligence and grade placement in conditioning achievement—although intelligence obviously ranks above grade placement. Neither is it possible to conclude (at least from data in this paper) that intelligence, as it is here measured (on the Terman Group Test of Mental Ability), is native rather than trained, or partly trained. But the study is none the less valuable in pointing which way the wind blows.

CHAPTER VI

GRADE PLACEMENT VERSUS MENTAL AGE AS A FACTOR IN SCHOOL ACHIEVEMENT

M. J. Van Wagenen
University of Minnesota, Minneapolis, Minnesota

Wherever mental tests have been given in elementary or high-school classes, wide variability in the mental age of the pupils has been the condition commonly found. It is not unusual to find in the same class pupils differing in mental age by as much as four years. Nor are the variations in attainment in any phase of school work usually discovered to be much less. Yet to many school people these conditions seem far less disturbing than wide variations in chronological ages. The latter condition results, of course, whenever gifted pupils are allowed to skip grades or backward pupils are required to repeat grades. Where ability grouping is feasible, this condition, however, can be largely eliminated. But the fact that many backward pupils are still likely to be found in grades in advance of their mental ages, while many gifted pupils are discovered in grades far below where their mental ages would place them, raises the question as to the comparative importance of grade placement and mental age in school attainment. Gifted pupils not infrequently do a poor quality of work in comparison to their capacity and the backward, as a rule, surpass them in relation to capacity. Is it, then, superior attainment of the less gifted, and low attainment of the gifted pupils, that accounts for the wide variability in attainment and mental ages in the classes?

Comparison of the mid-attainments of several hundred boys of corresponding mental ages in the sixth and eighth grades of a large school system affords a very definite answer to this question, and also indicates the relative importance of grade placement in relation to attainment in contrast with capacity to achieve as measured by mental tests.

In Table I are given the mid-scores made in eight achievement scales[1] by all the sixth-grade boys with mental ages ranging from 12 years, 3 months, to 15 years, 9 months. The same tendencies are also evident among the mid-scores of the girls, but the sex differences are too great in several instances to warrant combining the original scores for comparative purposes. For this reason, the scores of only the 333 sixth-grade boys and 486 eighth-grade boys have been used as a basis for attacking the problems suggested. To obtain more reliable mid-scores, the pupils were sectioned into groups each covering a range of three months in mental age, even though this procedure slightly minimized the differences among the mid-scores.

A glance down any column in Table I shows a decided increase in the mid-scores corresponding to the consecutive mental-age intervals for every phase of instruction measured. There are reversions, of course, but these should be expected as a matter of chance with so few cases in the consecutive intervals. The general tendency, however, is too marked to leave any doubt. While some backward pupils were undoubtedly promoted more rapidly than their mental ages would suggest and some gifted pupils were retarded because of failure to achieve, this was certainly not true of the group as a whole. The groups of pupils with the lower mental ages are the groups with the lower mid-scores. Likewise, the higher mid-scores were made by the pupils with the higher mental ages. A comparison of the mid-scores made by the sixth-grade boys with mental ages between 12 years, 3 months, and 12 years, 5 months, inclusive, with those made by the boys with mental ages of 15 years, 3 months, to 15 years, 5 months, inclusive, shows some striking differences. In spelling the two groups vary by more

---

[1] The mental tests and educational scales used in making the measurements were those of the series that have been rather widely used in recent achievement surveys; namely,

The Terman Group Test of Mental Ability, Examination A.
Spelling Scale A of the Van Wagenen Spelling Scales.
The Thorndike-McCall Reading Scale, Form 7.
The Van Wagenen American History Information Scale R.
The Van Wagenen American History Thought Scale R.
The Posey-Van Wagenen Geography Information Scale R.
The Posey-Van Wagenen Geography Thought Scales R and S.
The Woody Arithmetic Scale R (Van Wagenen revision) for the fundamental operations.
The Buckingham Scale for Arithmetic Problems, Form II, with the C score method of scoring.

TABLE I.—MID-SCORES OF 335 SIXTH-GRADE PUPILS CLASSIFIED BY MENTAL-AGE INTERVALS OF THREE-MONTHS

Mental Ages	Spelling	Reading	History Information	History Thought	Geography Information	Geography Thought	Arithmetic Fundamentals	Arithmetic Problems	No.
12_3	50.5	48	56	68	67	68	71	61	50
12_6	51.5	48	54.5	69	66.5	67	72	62	56
12_9	52	48	55.5	72	69	69	78	66	56
13_0	54	53	58.5	75.5	71	71	73	65	37
13_3	51.5	51	57.5	72.5	72	70.5	76	69	33
13_6	55	55	63	77	74	73	75	69	23
13_9	55	55	61	79	75	74	77	64	21
14_0	54	55.5	63.5	79.5	75	76	77	72	19
14_3	58	53	64.5	75.5	73	75	82	71	8
14_6	58	55	63	83.5	75.5	76.5	92	75	7
14_9	56.5	59.5	67.5	88	79	76.5	79	72	8
15_0	58	61	64.5	81.5	75.5	78	93	82	4
15_3	58	59.5	68.5	71.5	77.5	79.5	83.5	74	4
15_6	56	61	70.5	88.5	83.5	78.5	82.5	82	4
15_9	56.5	65	70.5	87	79.5	78	83	73	5

TABLE II.—MID-SCORES OF 486 EIGHTH-GRADE PUPILS CLASSIFIED BY MENTAL-AGE INTERVALS OF THREE-MONTHS

Mental Ages	Spelling	Reading	History Information	History Thought	Geography Information	Geography Thought	Arithmetic Fundamentals	Arithmetic Problems	No.
12_3	53.5	52	66	72.5	68	65.5	59.5	66	10
12_6	58	54	66.5	71.5	64.5	67	74	67.5	14
12_9	57.5	54	74	78	68	71.5	79.5	77.5	22
13–0	56	54	71	77	69.5	70	78	73	20
13_3	59	55	72.5	77.5	72.5	70.5	74	74	27
13_6	57.5	55	71.5	80.5	71	71.5	76	75	36
13_9	59.5	57	78	83	74.5	73.5	77.5	77.5	36
14_0	57	57	73	83.5	72.5	73	82	82	58
14_3	58.5	59	76	85	74	75	80	77	41
14_6	59	60	75.5	83	74.5	75.5	81	82	42
14_9	61	63	75.5	87.5	75.5	75.5	82	83	47
15_0	60.5	63	76.5	89	76.5	75.5	80	84	39
15_3	60.5	61	77	89.5	77	79	83.5	83	46
15_6	64	63	77	88.5	77	79	86.5	85	24
15_9	62	65	80	90.5	78.5	81	85.5	87	24

than a year, and in the thought phase of geography and the fundamental operations of arithmetic, by nearly two years. In the information phase of both American history and geography they vary by nearly three years, while in reading and thought phase of both American history and arithmetic they vary by more than three years. Thus, the two grades differed more widely in three phases in instruction than in mental age.

An inspection of the columns of Table II, which contains similar data for the eighth-grade boys of the same school system, reveals exactly the same tendencies. In every phase of work measured except their ranges of information in American history, there is a difference of two grades or more in the mid-scores of the 12 year, 3 months-old group and those of the 15 year, 3 months-old group. Even in the information phase of American history there is a difference of nearly two grades, while in the thought phase there is a difference of three grades. In abilities to solve arithmetic problems the difference between the mid-scores of the two groups is even wider than the differences in mental ages. Very clearly, the gifted pupils are not held back because of low attainments, but in spite of high ones. Likewise, the backward pupils are pushed along more rapidly than their attainments would warrant. Both of these conditions may be advisable when the pupils are grouped into homogeneous classes. But one at once questions whether this tendency in the normal class may not account for the fact that the relative quality of work of the gifted pupils is commonly below the average for all pupils and much below that of the backward pupils. While this question can be definitely settled only on the basis of experimental evidence, further analysis of the data in Tables I and II will show the relations of grade placement and mental age to school attainment.

As one compares the corresponding columns in Tables I and II for boys of the same mental age, the mid-scores of the eighth-grade pupils are for most groups higher than those of the corresponding sixth-grade groups. To some extent, of course, this is to be accounted for by the fact that the pupils with the higher attainments for their mental ages are undoubtedly more likely to be promoted or allowed to skip grades. How much is due to this factor cannot be determined from these measurements. But even if this is responsible for only a small amount of the difference, grade placement cannot be a very important factor in school achievement. As one compares the corresponding columns, mid-score by mid-score, the differences are found to be small, except in the case of range of information in American history, a phase of instruction upon which much emphasis was placed in the eighth grade of this school system. In most instances it is the similarities, rather than the differences, that are striking.

In Table III a qualitative comparison is made between the differences in the attainments of sixth- and eighth-grade pupils of corresponding mental ages and the differences in attainments of pupils of the same grade but differing by two years in mental age. Each item is an average of from seven to fifteen differences, and hence is a reasonably reliable basis on which to evaluate the tendencies. In the case of spelling the average two-year mental-age differences in Column III are just as great as the average sixth- and eighth-grade differences. In ability to answer questions on what is read, as measured by the Thorndike-McCall reading scales, the average two-year mental-age differences are more than two and a half times as great as the average difference between sixth- and the eighth-grade pupils of the same mental age. The ratio is nearly as great in the thought phase of American history.

TABLE III.—COMPARISON OF TWO-YEAR AND TWO-GRADE DIFFERENCES

	Average of differences for each of the seven two-year intervals of mental age in Grade VI	Average of differences for each of the seven two-year intervals of mental age in GradeVIII	Average of differences for each of the fourteen two-year intervals of mental age in Grades VI and VIII	Average of differences between mid-scores of Grades VI and VIII for each of the fifteen mental age intervals
Spelling	4.5	3.5	4.0	3.9
Reading	8.0	7.7	7.9	3.0
History information	9.0	5.3	7.2	11.4
History thought	8.9	10.4	9.7	4.6
Geography information	7.0	6.4	6.7	−1.2*
Geography thought	7.0	7.3	7.1	−0.5*
Arithmetic fundamentals	10.4	8.6	9.5	−1.0*
Arithmetic problems	10.4	10.0	10.2	7.6

*The minus sign indicates that on the average pupils of the eighth grade were below those of the same mental age in the sixth grade, even though they had studied geography during the seventh grade and the fundamental operations of arithmetic during both the seventh and eighth grades.

Even in the information phase of American history, where the marked emphasis upon American history in the eighth grade would lead one to expect much greater grade differences than mental age differences, the average grade differences are less than sixty percent greater than the average two-year mental-age differences. Although the eighth-grade pupils apparently know less about either the information or thought phases of geography than do the sixth-grade pupils of the same mental ages—in spite of the fact that they studied it intensively throughout the seventh grade—the pupils of

either grade who differ by no more than two grades in mental age differ by more than a grade[2] in attainments in all four cases. Just as in geography, the sixth-grade pupils stood higher in the ability to do the fundamental operations of arithmetic than did the eighth-grade pupils of the same mental age. The difference between the two-year mental-age levels was more than a year and a half in attainment. This should not be interpreted to mean that pupils do not make any gains in geography and the fundamental operations of arithmetic from the sixth grade to the end of the eighth grade. They make very nearly normal gains, but in the eighth grade the same pupils will on the average be mentally two years older than in the sixth grade. It is the difference in mental age, rather than the difference in grade placement, that is evidently responsible for the differences in attainment. This does not imply that gains would not have been made without instruction, but it does suggest that mental age is far more important in determining school attainment than either the differences in instruction or the differences in content from the sixth grade up through the eighth grade. The pupil's intellect, present abilities, and ranges of information are evidently far more significant than his grade placement. It is possible that too high a grade placement in comparison with mental age may serve as a stimulus to the less gifted pupil. Of this there is no evidence. That too low a grade placement may fail to stimulate the gifted pupil is not at all improbable.

This conclusion may seem open to several criticisms. The mental-age intervals for the higher levels may be inaccurate. Even if this were true, it would not invalidate the results. It is far more likely that the probable errors of the individual scores tend to minimize mental-age differences more than inaccuracies of mental ages would tend to exaggerate them. Even the crude mental-test scores would demonstrate the tendencies.

Again, it may be objected that mental age is but another index of school achievement. It is possible that school work, and especially school attendance, may have some effect upon the mental-test scores. Yet, even were this demonstrated, it would not affect the validity of these results. Since pupils with the lower attainments

---

[2] In all phases of instruction except spelling and reading a difference of 6 points represents the normal gain during a grade. In spelling and reading the normal gain for a pupil of average intelligence is 4 points.

tend to have still lower mental ages and pupils with the high attainments still higher mental ages, the effect of school work upon mental age, apart from the general functioning of intellect, is probably slight. In view of the fact that the pupils with the higher mental ages in any grade are the youngest chronologically and have skipped, rather than repeated, earlier grades, and that the pupils of the eighth grade have, in addition, attended school two years longer, whatever effect school work may have had in raising the mental age of the more gifted pupils of the sixth grade is more than offset by their shorter period of attendance at school.

Still a third criticism may be cited: namely, that the scales used in measuring attainment test what the pupils acquire outside of school rather than what is acquired in school. This probably could not be said of spelling and almost surely not of the fundamental operations of arithmetic; yet in the former the gain is the same for the difference of two years of mental age as for the difference of two school grades. In the case of the fundamental operations of arithmetic the average difference for two years of mental age is exceeded by only two other average mental age differences, those of the thought phase of history and of arithmetic problems; yet pupils of the eighth grade have a lower attainment than pupils of the sixth grade of the same mental age. If the scales measured what is acquired outside of school, rather than within, surely the chronologically younger pupils of the sixth grade would not know more than pupils of the same mental ages in the eighth grade, even though the latter had studied a year longer in school. Even in the case of reading ability, the average difference in mental age for two years is almost exactly what one would expect as the result of two years' work in school: the difference between the sixth and eighth grade mid-scores is only one-half of the normal gain that should result from two years' work. Nor was reading at all neglected in the seventh or eighth grade of this school system. The evidence in favor of the greater importance of mental age as compared with grade placement, even though it cannot be expressed with any degree of exactness, nevertheless seems almost unquestionable. This does not mean, however, that the gifted pupil can be just as effectively taught in a lower grade as in a group of pupils similar in mental age. But it does imply that content of instruction should be adjusted to mental age and present abilities.

PREFATORY NOTE TO CHAPTER VII

At the request of the Yearbook Committe, Mr. McCall has presented in Chapter VII a brief summary of a study that had been reported in detail earlier. The reader may consider it as a contribution along the same general lines as the two preceding chapters, that is, as an attempt to measure the amount of alteration in school achievement that can be expected from alterations in the educational environment, particularly in connection with teaching or classroom methods.

In this particular case the environmental alteration was the segregation of bright pupils. Sixty-seven bright pupils who were taught for two years in classes with other bright children only were compared for achievement with 67 equally bright children who were taught for two years in classes with all sorts (psychologically speaking) of children.

The outcome points to a definite advantage in segregation, although the conditions under which the study was made were not such as to permit this conclusion to be established beyond cavil from the statistical point of view.

In the *Twenty-Third Yearbook, Part I,* in Section II, dealing with special studies of gifted children, will be found several reports of investigations dealing with this question of the effects of segregation upon the school attainments of intellectually superior children.—*Editor.*

CHAPTER VII

COMPARISON OF THE EDUCATIONAL PROGRESS OF BRIGHT PUPILS IN ACCELERATED AND IN REGULAR CLASSES

William A. McCall
Teachers College, Columbia University, New York City

Purpose

This study compares the achievement in standard educational tests of 67 bright children who were taught in classes with other bright children for two years in a New York City public school with 67 unsegregated bright pupils scattered in several other New York City schools whose mental ages and chronological ages matched those of the segregated pupils, both pupil by pupil and for the group as a whole. Thus, both groups were equally bright. They differed chiefly in that the pupils in the former group were segregated whereas those in the latter group were taught in classes with pupils of widely and typically varying brightness, mental age, and chronological age. The 67 children in the segregated group were about equally distributed over Grades III to VII inclusive, and those in the non-segregated group were similarly distributed.

Tests Employed

The test used to measure intelligence was the National Intelligence Test, Scale A. The educational tests were the Thorndike-McCall Reading Scale, the Woody-McCall Mixed Fundamentals in Arithmetic, and selected words from the Ayres Spelling Scale. All tests were administered at the end of two years of segregation. Reading age, arithmetical age, and spelling age for each pupil were combined to get the educational age, and this was averaged for each group.

Essential Results

The following represent the essential findings of the investigation. The mean educational age for the segregated group was 145.42. For the non-segregated group it was 138.28. The differ-

ence of 7.14 is in favor of segregation. The standard deviation of the difference was 4.00, which yields an experimental coefficient of 0.62. Interpreted in words, this means that the segregated group had gained an excess of 7.14 months in 2 years, and we can be 62 percent certain that such segregation is in general more favorable to the growth of the traits measured than is non-segregation.

Comments

In an ideal experiment, gain from initial to final measurements would be preferable to final status only, but this study grew out of records collected for another purpose. Also, in an ideal experiment, an effort would be made to control possible differences in teaching efficiency, supervisory efficiency, and curricular content. It is possible that all three of these variables were favorable to the segregated group, since the pupils were in an experimental school. The amount of influence, if any, of such variables upon the final difference is not known. It is improbable that these influences were sufficient to explain all the superiority of the segregated group, particularly since it was the avowed object of the school to diversify the curriculum rather than to push the children forward in the particular traits measured.

PREFATORY NOTE TO CHAPTER VIII

This chapter continues the line of argument of the preceding chapters, though directing attention to a different environmental factor. The question here studied is: Can it be shown that the achievement of pupils, measured after twelve years of public-school instruction, is measurably superior in school systems that spend the most money in operating their schools. The answer is 'No.'

The investigation shows that educational costs, measured as the ratio of total current expenditures to average daily attendance, have no correlation with the gross amount of school subject matter (as measured on the Iowa High-School Content Examination) learned during twelve years of schooling. Neither do expenditures show any correlation with the composite marks which graduates of the school systems in question are able to make during their first semester at college.

If studies carried through in the future corroborate the indications of this chapter, it will behoove our schools to take careful stock of their aims, their methods, and their results, and possibly to formulate new aims. If cheaply run schools are as effective as expensive schools, both in teaching pupils prescribed content and in preparing them to learn in the future, very important questions are raised. Should we spend less money for our schools than we now do? Perhaps no educator or social scientist would say we should, but nevertheless the school administrator can hardly avoid asking himself, when he reads this chapter, just what the educational gains are that are supposedly assured by increasing the per capita expenditure for the operation of the school.

CHAPTER VIII

EFFICIENCY OF TRAINING AS AFFECTED BY THE COST OF INSTRUCTION

T. C. HOLY
Ohio State University, Columbus, Ohio
and
G. M. RUCH
University of California, Berkeley, California

THE PROBLEM

This investigation was directed at the determination of the relation between the amount of money spent by school systems per pupil in average daily attendance and the effectiveness of instruction.[1]

THE MEASURES USED

Four different measures were used in the attempt to evaluate this problem, as follows:

(1) School costs; computed by the conventional formula, Total Current Expenditures[2] divided by Average Daily Attendance

(2) The Thorndike Intelligence Examination for High-School Graduates, Part I, Form B

(3) The Iowa High-School Content Examination,[3] Form A

(4) Composite marks of high-school graduates earned during the first semester of the year 1923-24 at the State University of Iowa

---

[1] Extracts from a larger study by T. C. Holy entitled: "Relation Between Cost Per A. D. A. and Quality of Scholarship in Iowa Public Schools," 1924. Unpublished to date. This study was made when both Professor Holy and Professor Ruch were connected with the State University of Iowa.

[2] "Total Current Expenditures" was defined as the money expended for all purposes other than capital outlay, interest on bonded indebtedness, and bond payments. It includes the salaries of superintendents, supervisors, teachers, janitors, and other help; tuition costs; texts and library books; instructional and janitorial supplies; transportation costs; fuel; rent; insurance; interest on temporary debt; and certain other items.

[3] This is a test with 80 minutes' working time, containing 400 items covering the more important subjects of the high-school curriculum. It shows a reliability well above 0.95 for the subjects studied.

The school costs were figured for the school years 1920-21 and 1921-22. Graduates of 26 Iowa high schools in the spring of 1923 and all freshmen entering the State University of Iowa during the fall of 1923 were included. The marks earned in all college subjects were weighted by the method of standard measures to eliminate, as far as possible, differences in departmental standards.

Both measures of efficiency of instruction (The Iowa High-School Content Examination and the composite marks) are admittedly but partially satisfactory criteria, but both must be accorded considerable weight in evaluating the results of public-school training. The Iowa High-School Content Examination measures *directly* the effects of high-school instruction and *indirectly* to some extent the effects of elementary instruction. College marks are valid for purposes of this study to the extent to which they reflect previous preparation, habits of work, etc.

The Pupils Studied

In all, 1796 pupils were studied, but the present report will be confined to pupils who completed all 8 grades of elementary and all 4 grades of high-school instruction in the *same* Iowa school system. The cost per pupil in average daily attendance can be computed rather exactly for these pupils who were educated throughout within a given school system.

Classification of Schools

Since school costs differ to some extent with the size of the school system, five classes of schools are considered, based upon number of pupils in average daily attendance, as follows:

Class I	1 — 199
Class II	200 — 399
Class III	400 — 799
Class IV	800 — 1699
Class V	1700 — above

The Results

Table I shows the correlations between the three measures used for the study. In explanation, the values in the column r_{12} are read: "the correlation between accomplishment as measured by

the Iowa High-School Content Examination and cost per pupil in average daily attendance." The values in the column head $r_{12.3}$ may be read: "The correlation between accomplishment and cost with the influence of intelligence eliminated."

TABLE I.—CORRELATIONS, WITH THEIR PROBABLE ERRORS, BETWEEN THREE MEASURES IN FIVE CLASSES OF SCHOOLS
(Data for high-school examination.)

Class of School	r_{12}	r_{13}	r_{23}	$r_{12.3}$	N
I......	−.036 ± .073	.67 ± .040	.099 ± .071	−.14	87
II......	−.15 ± .055	.66 ± .031	−.016 ± .056	−.19	142
III......	.067 ± .052	.67 ± .028	.17 ± .050	−.079	168
IV......	−.019 ± .050	.64 ± .029	−.13 ± .049	.086	179
V......	−.076 ± .054	.77 ± .022	−.055 ± .054	.052	159

The variables in this table are numbered as follows:
1. Iowa High-School Content Examination.
2. Cost.
3. Thorndike Intelligence Examination.

Table II presents similar coefficients of correlation where the first-semester college marks are used as measures of accomplishment instead of the Iowa High-School Content Examination.

TABLE II.—CORRELATIONS, WITH THEIR PROBABLE ERRORS, BETWEEN THREE MEASURES IN FIVE CLASSES OF SCHOOLS
(Data for college marks.)

Class of School	r_{12}	r_{13}	r_{23}	$r_{12.3}$	N
I......	−.063 ± .079	.54 ± .056	.017 ± .080	−.086	71
II......	.004 ± .064	.61 ± .044	.040 ± .063	−.025	112
III......	.11 ± .068	.53 ± .050	.11 ± .068	.059	95
IV......	−.15 ± .070	.46 ± .056	−.026 ± .071	−.16	89
V......	.14 ± .063	.59 ± .042	−.041 ± .071	.18	110

The variables in this table are numbered as follows:
1. Composite first semester college marks.
2. Cost.
3. Thorndike Intelligence Examination.

CONCLUSIONS

1. Tables I and II show no significant correlations between school costs and either of the two measures of effectiveness of instruction, whether zero order or partial correlations eliminating intelligence are considered.

2. Intelligence, as measured by the Thorndike test, is of far more importance in determining the effectiveness of training than is cost per pupil in average daily attendance (Table I).

3. So far as preparation for successful accomplishment in college is concerned (whether fundamental training, habits of study, or what), the cost of public-school education seems not to influence to any important extent attainment in institutions of higher learning (Table II).

4. The criteria of the effectiveness of training, although probably rather imperfect measures, would seem to possess enough value to permit effects of varying instructional costs to emerge if these were highly important influences in achievement.

TABLE III.—THE MEANS AND STANDARD DEVIATIONS OF THE MATERIAL USED IN TABLE I

Class	Means			Standard Deviations		
	I.H.C.E.	Cost	Thorndike	I.H.C.E.	Cost	Thorndike
1.	174.4	$115.82	84.6	33.9	$ 47.80	19.4
2.	174.6	155.05	92.4	46.9	44.20	21.2
3.	180.4	100.11	95.5	46.5	18.15	16.9
4.	191.8	77.95	97.0	54.2	14.82	18.7
5.	189.9	94.11	99.7	58.5	11.82	21.3

TABLE IV.—THE MEANS AND STANDARD DEVIATIONS OF THE MATERIAL USED IN TABLE II

Class	Means			Standard Deviations		
	Marks	Cost	Thorndike	Marks	Cost	Thorndike
1.	12.0	$130.16	87.1	29.0	$ 34.40	20.0
2.	31.0	137.25	93.0	13.6	32.72	21.6
3.	32.5	100.16	101.42	12.8	24.45	17.2
4.	29.2	82.23	92.6	12.5	15.54	19.1
5.	36.0	89.82	102.9	13.6	11.04	20.3

PREFATORY NOTE TO CHAPTER IX

The conclusion of this chapter seems quite clearly established—"that inferior students study somewhat more hours, [than do brighter students] but that this slight increase in duration of study does not compensate at all for their lack of ability." The correlation between the number of hours per week which a group of college students spent in studying had a slight negative correlation with the grade-points they earned during their first college semester. This correlation became zero when proper discount was made for their composite intelligence scores on four examinations.

It may seem surprising that, even for students of equal mental ability, time spent in study fails to differentiate them according to success in class work. Certainly, we have all known college students who "flunked" because they did not apply themselves, and others, not necessarily of the highest mental caliber, who graduated with honors by dint of hard work. Possibly the zero partial correlation can be partly accounted for by one or more of the following conditions:

1. The unreliability of college marks.

2. Unrepresentativeness of the 'study time' measures, owing to the fact that the students kept records of study only during one week.

3. The tendency of many students to study only enough to 'get by.'

4. Differences in difficulty of various courses, thus necessitating different amounts of study for different classes to insure the same class marks.

This investigation does not prove that inferior students *could* not compensate their inferior ability by increased time spent in study; but it does show that, as a group at least, they *do* not do it.

CHAPTER IX

ACHIEVEMENT AS AFFECTED BY AMOUNT OF TIME SPENT IN STUDY[1]

LONZO JONES
Graceland College, Lamoni, Iowa
AND
G. M. RUCH
University of California, Berkeley, California

INTRODUCTION

It is a common assumption that the amount of time spent in study is an important factor in determining accomplishment. This is very probably true within certain limits if other factors are taken into account. One variable in the situation would seem to be the fact that the *amount of time* spent in study is not a very perfect measure of the *effectiveness* of study. Another factor would seem to be differences in the abilities of the students—whether these be differences in original endowment or in previous achievement. In the third place, there are probably definite limits within which time spent in study influences learning.

A practical difficulty in recording time spent in study for experimental investigation arises from the fact that much must be left to the honesty of the students who keep their study records. This limitation is inherent in the data reported here, except that two facts guarantee that dishonesty was kept at a minimum, viz.; (1) confidence and understanding on the part of the students was established by means of preliminary personal interviews and explanations, and (2) subsidiary investigations showed substantial agreements in the records kept by students on themselves and those kept for the same students by responsible persons in charge of their living groups.

THE METHOD

Two groups of college freshmen will be considered here; first, 130 unselected students, and second, 150 freshmen so selected as to provide three groups of 50 each that differed widely in mental ability.

[1] This investigation was conducted when both of the authors were connected with the State University of Iowa.

All these students kept an hour-by-hour time chart for one typical school week. A week was chosen that would be free from any unusual disturbing factors in study, such as major football or other athletic events, a vacation immediately preceding or following, or a final examination period, etc.

A three-hour mental examination had previously been given to all students included in the study. This consisted of four tests:

Thorndike Intelligence Examination, Part I, Form E
Thorndike Intelligence Examination, Part I, Form O
The Morgan Mental Test
The Iowa Comprehension Test

The results of these four tests were weighted and pooled and then expressed as percentile ranks for a group of about 1200 State University of Iowa freshmen.

The Results

Table I shows the average number of hours spent in study for the week selected and the average grade-points earned at the end of the first semester of the year 1923-24—the semester which included the week during which the time records were obtained.

Table I.—Time Spent in Study and Grade-Points Earned by Three Groups of Freshmen at Iowa

Group	Average No. Hours Spent in Study for One Week	Average Grade-Points Earned During the Semester
A. 50 superior men (Top 10 percent in the mental examination)..........	29.1 ± .72	46.1 ± .99
B. 50 average men (Between the 43d and the 56th percentile in the mental examination)....................	33.7 ± .92	23.8 ± 1.25
C. 50 inferior men (Bottom 15 percent in the mental examination).......	34.3 ± .87	7.0 ± 1.44

Table I makes it appear that the dull students study somewhat more in terms of hours than do the superior ones; the difference, however, is not very marked. There is, of course, the possibility that the records were 'padded,' but this disturbance would have

to be considerably greater for one group than another to change significantly the complexion of Table I. The data show much wider differences in grades earned by the three ability groups than differences in time devoted to studying. In spite of the apparently greater studiousness of the less able students, they accomplish far less in the judgment of their instructors. Two interpretations are possible: (1) the inferior student really does study more because of necessity, but even then fails to overcome the handicaps of relative lack of ability; or (2) the inferior student is the more

TABLE II.—TIME SPENT IN STUDY AND GRADE-POINTS EARNED BY 130 UNSELECTED FRESHMEN AT IOWA

	Grade-Points Earned	Percentile Ranks on the Mental Examination	Hours Spent in Study for the Week in Question
Average	26.2	52.7	31.4
Standard Deviation	16.2	30.6	9.0
Range	−20 to 59	1 to 99	13 to 58

Correlations between (1) grade-points, (2) percentile rank, and (3) hours of study.

$r_{12} = .69 \pm .031$ $r_{12.3} = .66$

$r_{13} = -.28 \pm .055$ $r_{23.1} = -.32$

$r_{23} = -.41 \pm .049$ $r_{13.2} = .0044$

$R1.23$ $.69$

dishonest about his records. It is our opinion, based upon observation of these students over a period of a year, that the first interpretation more nearly pictures the real situation. In many cases it was possible to confirm the substantial accuracy of the time records submitted.

Table II shows similar results for 130 unselected university freshmen for the same year and class, together with the correlations between grade-points, ability, and hours spent in study for this typical group of freshmen. It will be noted that mental ability shows a correlation of .69 with grade-points earned and that hours spent in study shows a slight, but significant, negative correlation (−.41) with achievement. However, when the factor of mental ability is eliminated through partial correlations, the former relationship remains constant, while study time disappears as a factor operating over and above differences in ability.

Conclusion

If these data are to be taken as typical, the conclusion is forced upon us that inferior students study somewhat more hours, but that this slight increase in duration of study does not compensate at all for their lack of ability. Study time, as voluntarily scheduled by the average student, would appear to be a factor in achievement of much less importance than ability as measured by the mental examinations employed. The amount of study seems to be dictated by the need of the individual student to maintain a quality of accomplishment which will permit him to remain in college, and this need is inversely proportional to his ability.

To avoid misinterpretation, let us note that these results are based upon the study time as *voluntarily* scheduled by average freshman students. It need not be concluded that additional time spent in study is futile, but merely that the slight increase that was actually reported by the less gifted students is not a decisive factor. The problem may be raised, but is not answered here, whether the time spent in study by the less able students could be increased to such an amount as to produce a marked increase in grade-points earned by them.

PREFATORY NOTE TO CHAPTER X

The chapters in this second section, Nos. X, XI, and XIII, deal with the relative influence of heredity and environment upon certain special traits, in contrast with the preceding chapters which were concerned with the relative influence of heredity and environment upon general school performance. In addition to these three chapters, the reader may consult Chapter XV for a summary of the literature dealing with the relation of heredity and environment to another special trait, musicality.

Comparatively few studies have been published dealing with this influence of nature and nurture upon specific mental traits. Anderson's chapter on mechanical ability in relation to environment is, we believe, the first study of its kind.

The results of his study indicate that little or no influence can be attributed to the mechanical activities of a boy's father, the tools which the boy's father owns, or the cultural status of the home, upon the boy's mechanical ability when this is judged by ratings upon actual projects made in school shops, tests of 'shop information' at the close of shop courses, and scores on eight standard mechanical tests. The absence of any measurable influence from these environmental factors leads to the surmise that the undoubted differences that exist in mechanical ability are mainly innate.

CHAPTER X

THE RELATIONSHIPS OF CERTAIN ENVIRONMENTAL FACTORS TO MEASURES OF MECHANICAL ABILITY[1]

L. Dewey Anderson
University of Minnesota, Minneapolis, Minnesota

One of the problems in a research dealing with mechanical ability which has been in progress since 1923 at the University of Minnesota was to determine the relationship of environmental conditions to mechanical ability. In the effort to get an inventory of the former the particular-item and the rating-scale methods were studied to determine their relative merits for this purpose. It was finally decided to use the particular-item method, but to have enough items so that scores might be secured for different specific aspects of the environment.

A preliminary schedule of 500 questions was first made and used in interviewing the parents of ten boys. The social worker who made the calls, found that many parents were reluctant to give certain information and also that the mother, who was usually interviewed, did not know the names of the mechanical tools owned by her husband or son. In order to remedy these difficulties, the questionnaire was shortened and divided into two parts—Part I (Sections I, II, III, IV, and V) to be filled out by the social worker on her call, and Part II (Sections V and VII) to be filled out and mailed by the parents. The questionnaire was so arranged that items dealing with given aspects of the environment were grouped in a section by themselves. The answers were recorded by inserting in the parenthesis to the left of the items a "1" for possession (or an affirmative answer) and "0" for non-possession (or a negative answer).

A trained social worker[2] secured the information in Part I by visiting the homes and questioning the parents of the boys in the

[1] This study was a part of an investigation of mechanical ability directed by R. M. Elliot and D. G. Paterson, of the University of Minnesota. The investigation was sponsored by the Committee on Human Migrations of the National Research Council. Funds were supplied by the Laura Spelman Rockefeller Foundation.

[2] Mrs. G. B. Braithwaite, trained in social work at Simmons College, with two years of experience as visiting teacher in the Rochester, New York, public schools.

experimental group that was next studied. This experimental group consisted of 150 boys enrolled at the Jordan Junior High School, Minneapolis, for whom we had scores in eight mechanical ability tests and reliable objective indices of success in mechanical shop courses.

In interviewing the mothers of the ten boys in the preliminary, or trial, group, the social worker felt that the progress of the interviews was hampered because she did not know the boys. In order that she might become acquainted with the boys, she called each boy out of the classroom for the ostensible purpose of giving him a short psychological test. At this meeting she explained to him that she would visit his mother within a few days. These acquaintanceships proved very valuable, as they gave the field worker personal 'leads' for conversation with the mother when it was evident that a boy was not interested in mechanical things. At the close of the interview the social worker left the second part of the questionnaire with the mother, and asked her to fill it out and mail it. The time of the individual call varied from twenty minutes to three hours. Before the social worker made a call, the principal of the school in which the boys were enrolled sent the following letter to the parents.

Mr. and Mrs. ....

Minneapolis, Minnesota

Dear Mr. and Mrs......................

Last year your son ....took some tests in the Jordan School for the purpose of determining his mechanical ability and of aiding us more definitely to arrange courses of study to meet the needs of the boys of the school. In order to have more information upon which to work, Mrs. .... will visit your home in the next few days, asking for some information regarding your boy and his out-of-school activities.

We will appreciate it very much if you will give Mrs. .... as much assistance as possible in securing the information that we need. Any information given will, of course, be confidential in our school office; so you need not hesitate to talk very freely with her. I thank you for this coöperation.

Sincerely,

............................Principal
Jordan Junior High School

It is probable that the coöperation of the principal in sending the letters and in giving a talk to the boys was mainly responsible for the excellent results. The parents of 143 of the 150 boys were visited, and 135 of the 143 families returned the second part of the questionnaire.

After the data were collected, the items were arranged in seven classifications, and then the number of affirmative responses was added to get a score for each. The number of items in each classification is given in the last column of Table I.

The items in each section of the questionnaire are given below. Preceding each item is given the percent possessing (or the percent of affirmative responses).

QUESTIONNAIRE ON CULTURAL AND MECHANICAL ENVIRONMENT

(Rearranged according to classifications used in statistical analysis)

The percent of families 'possessing' an item is placed in the left hand column. Items are arranged in order according to the frequency of possession.

I. Cultural Status of Family

Percent Possessing	*Name of Item*
(97)	Lighting—electric
(96)	Electric iron
(95)	Sanitation—toilet
(95)	Bath tub
(75)	Homes having six rooms or more
(72)	Furnace
(69)	Radio
(68)	Floor coverings—polished, expensive rug
(64)	Kitchen floor—painted, linoleum
(63)	Phonograph
(63)	Telephone
(61)	Hot water system
(57)	Automobile other than truck owned by the family
(57)	Piano—upright, grand
(54)	Living room appraisal—excellent or good
(53)	Lamps—reading or floor
(45)	Electric washer
(44)	Electric vacuum sweeper
(44)	Dining room appraisal—excellent or good
(35)	Electric toaster
(21)	Purchased radio
(13)	Electric sewing machine
(5)	Typewriter
(1)	Electric percolator
(1)	Radio enclosed in cabinet
(0)	Electric stove
(0)	Artificial refrigeration
(0)	Fireless cooker
(0)	Steam pressure cooker

II. Literary Interests

Percent Possessing	Name of Item
(88)	Boy has library card
(56)	50 or more books in home
(45)	Number of magazines subscribed for—7 or more
(45)	Number of newspapers subscribed for—2 or more
(41)	Boy reads 6 or more books a month
(32)	Education of father—above 8th grade
(30)	Education of mother—above 8th grade

III. Recreational Interests of Boy

(93) Can spin tops
(91) Can ice skate
(91) Can play marbles
(87) Can fly kite
(85) Can swim
(83) Can ride a bicycle
(80) Can play football
(79) Can play baseball
(71) Can walk on stilts
(68) Takes care of pets
(68) Wanted a tool chest after taking manual training
(64) Has had a "Meccano" or "Erector" set
(58) Knows how to row a boat
(57) Wanted a tool chest before taking manual training
(57) Knows how to play basketball
(53) Knows how to play croquet
(50) Knows how to wrestle
(48) Knows how to play shinny
(23) Knows how to roller skate
(1) Knows how to play pool
(1) Knows how to bowl

IV. Mechanical Operations of Father

(83) Put on screens
(69) Replaced washers in faucets
(62) Replaced electrical fuses
(61) Painted houses
(60) Painted or varnished woodwork or floors
(59) Puttied around windows
(59) Fixed window shades
(59) Put in window panes
(56) Tightened springs in window shades
(54) Replaced light sockets
(53) Hung window blinds, screen doors or storm doors
(52) Repaired electrical circuits
(51) Repaired automobile engines
(51) Put up shelves
(50) Sharpened scissors and knives
(50) Repaired plumbing
(43) Soldered
(42) Cleaned out flues in stoves or furnaces
(37) Planed doors to fit
(35) Sharpened carpenter tools
(33) Fixed leaks in kettles
(31) Half-soled shoes
(26) Shingled or repaired a roof
(24) Made a radio set
(23) Riveted straps
(23) Sharpened saws
(21) Mended screens
(21) Built a garage
(19) Made an ironing board

Percent Possessing	Name of Item
(18)	Made a drain board
(18)	Repaired sewing machine
(16)	Made a chest of drawers
(16)	Replaced screening in doors or screens
(15)	Repaired furniture
(15)	Tinted or kalsomined walls
(15)	Repaired vacuum sweeper
(14)	Sharpened lawn mowers
(13)	Made trellises for vines
(13)	Built coal sheds
(13)	Made a kitchen cabinet
(12)	Papered walls
(11)	Threaded pipes
(11)	Made a bread board
(11)	Put down carpet or oilcloth
(10)	Built chicken coops
(10)	Built chicken roosts
(10)	Put pole in clothes closet
(8)	Mended door stoops
(7)	Made a footstool
(6)	Made a book case
(5)	Made a taboret
(5)	Made a cedar chest or box
(2)	Made bookends or a book-rack
(2)	Made a necktie rack
(1)	Put in water backs in stove

V. Mechanical Operations of Son

(68) Put on screens
(59) Replaced fuses
(53) Made a radio set
(52) Painted or varnished woodwork or floors
(52) Repaired engine in automobile
(52) Repaired light sockets
(49) Repaired electrical circuits
(42) Painted houses
(35) Replaced washers in faucets
(33) Soldered
(33) Sharpened scissors and knives
(30) Put up shelves
(28) Cleaned out flues in stoves or furnaces
(28) Made a necktie rack
(28) Fixed leaks in kettles
(22) Sharpened carpenter tools
(22) Puttied around windows
(22) Riveted straps
(21) Tightened springs in window shades
(14) Repaired sewing machines
(13) Hung window blinds, screens, or stormdoors
(12) Repaired plumbing
(11) Mended screens
(10) Made a footstool
(10) Shingled or repaired a roof
(9) Repaired vacuum sweepers
(8) Built a garage
(8) Made book ends or book-racks
(8) Replaced screen wire in doors or screens
(8) Repaired furniture
(7) Built coal sheds
(7) Put down carpets or oilcloth
(7) Tinted or kalsomined walls
(7) Made trellises for vines
(6) Put pole in clothes closet
(6) Planed doors down to fit
(6) Made bread boards
(6) Half-soled shoes
(5) Built chicken roosts
(5) Mended door stoops
(5) Made a taboret
(4) Built chicken coops
(4) Made a cedar chest or box
(4) Threaded pipes
(4) Sharpened lawn mowers
(3) Papered walls

Percent Possessing	Name of Item	Percent Possessing	Name of Item
(2)	Made a kitchen cabinet	(1)	Made a chest of drawers
(2)	Sharpened saws	(1)	Made drain boards
(2)	Made an ironing board	(1)	Made a bookcase

VI. Tools Owned by Father

(89) Screwdriver
(88) Pliers
(86) Ax
(84) Claw hammer
(83) File
(83) Putty-knife
(80) Monkey wrench
(76) Chisels
(75) Carpenter's rule
(71) Whetstone
(69) Pipe wrench
(67) Brace
(67) Ordinary hatchet
(66) Glass cutter
(66) Buck saw
(64) Cross cut saw
(57) Plain soldering iron
(56) Level
(55) Nail punches
(54) Small square
(54) Block plane
(54) Tinner's snips
(53) Machinist's hammer
(53) Awl
(50) Steel square
(48) Shoemaker's stand
(48) Rip saw
(47) Hack saw
(44) Work bench
(44) Keyhole saw
(44) Leather punch
(43) Dividers
(41) Vise
(41) Alligator wrench
(41) Compass
(41) Jack plane
(39) Emery wheel
(38) Miter square
(37) "S" wrench
(33) Yankee drill
(31) Smoothing plane
(31) Shoemaker's hammer
(31) Countersink bit
(29) Saw set
(28) Marking gauge
(28) Drawknife
(27) Miter box for saw
(27) Outside calipers
(25) Iron clamps
(25) Shoemaker's nippers
(21) Cabinet scraper
(21) Ratchet drill
(20) Blow torch
(20) Inside calipers
(20) Shingling hammer
(19) Riveting hammer
(19) Reamer
(19) Rivet set
(19) Plumbing force pump
(18) Coping saw
(15) Grooving plane
(13) Wood clamp
(8) Electric soldering iron
(8) Threading machine

VII. Tools Owned by Son

(46) Pliers
(40) Screwdriver
(35) Claw hammer
(31) Compass
(30) File
(25) Hatchet
(25) Chisel
(25) Axe

Percent Possessing	*Name of Item*
(22)	Carpenter rule
(21)	Monkey wrench
(18)	Soldering iron (plain)
(17)	Cross-cut saw
(16)	Glass cutter
(15)	Bits
(15)	Whetstone
(15)	Brace
(12)	Nail punch
(10)	Block plane
(9)	Vise
(9)	Machinist hammer
(9)	Awl
(9)	Buck saw
(9)	Rip saw
(9)	Small square
(9)	Steel square
(9)	Pipe wrench
(7)	Dividers
(7)	Leather punch
(7)	Tinner's snips
(5)	Drawknife
(5)	Emery wheel
(5)	Level
(5)	Jack plane
(5)	Coping saw
(5)	Keyhole saw

Percent Possessing	*Name of Item*
(5)	Alligator wrench
(5)	Marking gauge
(5)	Iron clamps
(4)	Hack saw
(4)	Miter box for saw
(3)	Riveting hammer
(3)	Cabinet scraper
(3)	Shingling hatchet
(3)	Shoemaker's peg snippers
(3)	Shoemaker's stand
(3)	Wood clamp
(2)	"S" wrench
(2)	Saw set
(2)	Shoemaker's hammer
(2)	Smoothing plane
(2)	Inside calipers
(2)	Outside calipers
(2)	Plumbing force pump or cup
(2)	Electric soldering iron
(2)	Yankee drill
(1)	Blow torch
(1)	Ratchet drill
(1)	Rivet set
(0)	Grooving plane
(0)	Reamer
(0)	Miter square
(0)	Threading machine

The reliability coefficients (see Table I) of the seven classifications were secured by correlating the sum of affirmative responses in the odd-numbered items with the sum on the even-numbered items and applying the Brown-Spearman reliability formula.

Only one of the seven coefficients, Literary Interests, which contained only seven items, was so unreliable (.47) as to make its use questionable in further statistical treatment. The reliabilities for the other six classifications indicate that relatively the same index in each aspect of a home would be obtainable from a second independent set of items.

One of the most important questions concerning the use of environmental questionnaires is: "Are the variabilities of the scores large enough to insure the heterogeneity of the group in these aspects?"

TABLE I.—INTERCORRELATIONS OF THE SEVEN SOCIO-ECONOMIC CLASSIFICATIONS N—100

	Economic and Cultural			Mechanical				Reliability Coefficient	Number of Items in Class
	Cultural Status of Family	Literary Interests of Son	Recreational Interests of Son	Mechanical Operations of Father	Mechanical Operations of Son	Tools Owned by Father	Tools Owned by Son		
Cultural and Economic:									
Cultural status of family	....	.33	.22	.01	.01	−.01	.06	.90	30
Literary interests of son	.33	....	.19	.15	.11	.06	−.01	.47	7
Recreational interests of son	.22	.19	....	.16	.39	.08	.17	.73	21
Mechanical:									
Mechanical operations of father	.01	.15	.16	....	.68	.49	.04	.94	54
Mechanical operations of son	.01	.11	.39	.68	....	.20	.06	.94	54
Tools owned by father	−.01	.06	.03	.49	.20	....	−.03	.94	65
Tools owned by son	.06	−.01	.17	.04	.06	−.03	....	.89	65

This question of variability is answered in part by an inspection of the distributions of the scores which are shown graphically in Figure I.

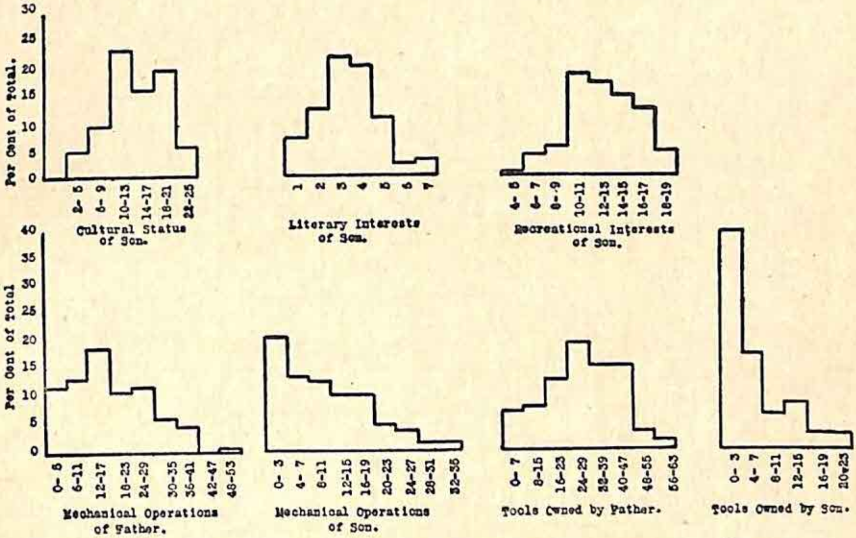


FIGURE I.—DISTRIBUTION OF SCORES IN SEVEN ENVIRONMENTAL MEASURES

The distributions are fairly normal in all but three sections, Son's Mechanical Operations, Father's Mechanical Operations, and Tools Owned by Son. In all cases except Tools Owned by Son, the scores are distributed throughout almost the entire possible range. For example, in Cultural Status the actual range is 2-26 and the possible range is 0-30. These two facts, wide range and normal distribution of scores within the ranges, indicate that the test group was a fairly heterogeneous one in respect to all the environmental aspects except Son's Mechanical Operations and Tools Owned by Son.

The intercorrelations of these environmental measures are given in Table I. If we consider only the classifications relating to the mechanical environment, the average intercorrelation is .24. The measures are somewhat related to each other—particularly is Mechanical Operations of Father related to the other items. If we eliminate Tools Owned by Son, we get an average intercorrelation of .46 for three other measures of mechanical environment—Mechanical Operations of Father, Mechanical Operations of Son, Tools

Owned by Father. The average intercorrelation of Cultural Status, Literary Interests, and Recreational Interests (economic and cultural factors) with the mechanical classifications (.10) indicates that more than one phase of environment is being measured by our questionnaire. The average intercorrelation of the three non-mechanical (*i.e.*, economic and cultural) aspects of environment is .25. This indicates a somewhat higher inter-relationship among these non-mechanical phases of the environment than exists between the non-mechanical and the mechanical phases.

These intercorrelations of economic and cultural status are not high enough to indicate that the non-mechanical classifications are measuring just one aspect of the environment. They only indicate that the non-mechanical classifications are measuring something of a different nature than the mechanical classifications and that the mechanical classifications are in turn measuring related, but rather specific, factors which we concluded from their apparent nature to be measures of mechanical environment.

In analyzing this table we found some evidence that the possession of material wealth, as indicated by Cultural Status and Literary Interests, was somewhat detrimental to mechanical operations and the possession of tools. A family with fair economic means probably hires somebody to do much of the repairing and odd jobs about the house. Therefore, it is not necessary for the father and son to do this work and consequently the son has more time left to read and play. The positive correlations (.22) between Recreational Interests and Cultural Status and (.33) between Literary Interests and Cultural Status and the lack of positive relationships between Cultural Status and the several measures of the mechanical environment indicate the probable truth of this statement.

To determine the relation of these environmental factors to mechanical ability, the seven aspects of the environment were correlated with two mechanical shop criteria, or scores in mechanical success. The shop success scores indicated (1) success in manipulation of tools and materials, and (2) amount of information acquired about tools and materials.

In obtaining the measures of manipulative skill, the projects made by the boys were subjected to objective measurement on many

well-defined and specific parts. These measurements were combined to give a general measure of manipulative success in the type of shop work covered by junior-high-school courses in sheet metal, woodwork, electricity, printing, and mechanical drawing. The amount of information acquired in each of these shops was measured by multiple-choice information tests given at the end of the course, the scores of which were combined to give a general measure of 'shop information.' The correlations between the environmental aspects and the shop-success measures are given in Table II.

TABLE II.—THE CORRELATIONS OF THE SEVEN ENVIRONMENTAL FACTORS WITH THE MEASURES OF MANIPULATIVE AND OF INFORMATIONAL STANDINGS IN MECHANICAL LINES (100 CASES)

Factor	Correlation with Manipulative Criterion	Correlation with Informational Criterion
Cultural status of family	−.07	.01
Literary interests of son	.13	.22
Recreational interests of son	.01	.12
Mechanical operations of father	.09	.24
Mechanical operations of son	.30	.35
Tools owned by father	.01	.11
Tools owned by son	.10	−.02

All of these correlations are low. The measures of mechanical environment show slightly higher relationship with mechanical ability than do those of non-mechanical environment, but are too low to warrant the conclusion that mechanical shop success is substantially related to those aspects of the environment measured by our questionnaire.

The correlations of the mechanical test scores on the individual tests of our mechanical prognosis battery also show low relationships with the measured environmental factor. These are given in Table III.

The low relationships in Tables II and III are not caused by low reliability of the criteria or the prognosis tests, since these all have high reliability. The reliability of the manipulative criterion is .76. We do not know the reliability of the informational criterion, but the reliabilities of the separate tests composing it range from .71 to .84. The reliabilities of the predictive tests are somewhat higher (see Table III).

TABLE III.—THE CORRELATIONS OF THE SEVEN ENVIRONMENTAL FACTORS WITH SCORES IN EIGHT MECHANICAL ABILITY TESTS AND A TEST OF GENERAL INTELLIGENCE

Environmental Factor	Minnesota Assembly	Paper Formboard	Spatial Relations	Packing Blocks	Card Sorting	Stenquist Picture I	Stenquist Picture II	Steadiness (9-hole test)	Intelligence, Otis, I.Q.*
Cultural status of family	−.14	.14	.02	−.04	.01	.01	.05	−.19	.14
Literary interests of son	.00	.25	.12	−.09	.03	−.08	.12	−.14	.36
Recreational interests of son	.23	.12	.05	−.05	−.05	.04	.17	.07	.09
Mechanical operations of father	.18	.23	.09	−.05	−.16	.10	.08	−.03	.09
Mechanical operations of son	.40	.24	.22	.00	−.12	.24	.19	.00	.10
Tools owned by father	.03	.05	.00	−.07	−.19	−.05	−.08	−.02	−.05
Tools owned by son	.14	.01	−.04	−.09	.05	−.06	−.07	.18	−.05
Reliability coefficients	.94	.90	.84	.77	.90	.89	.90	.93	

*The range of intelligence quotients was restricted by reason of the fact that the cases were all in the seventh grade at the time of testing.

There is some question concerning the effect of the form of the distributions of the three factors, Tools Owned by Son, Father's Mechanical Operations, and Son's Mechanical Operations, on the size of correlations with the scores of shop success. The correlations of Son's Mechanical Operations with the Manipulation Criterion and the Information Criterion were, respectively, .30 and .35. An inspection of the correlation data sheets did not show that an appreciable increase in these correlations would result if the forms of the distributions were corrected. A more exact method might raise the figures somewhat, but it is doubtful if there would be very marked increases.

The correlations of the environmental classifications and the Otis Group Test (Higher Examination, Form A) intelligence quotients are in the last column of Table III. Literary Interests of the Son shows a comparatively high correlation (.36) with the I.Q., whereas the mechanical classifications show very low correlations with it; two of them are positive and two negative.

Our results, then, indicate that there is practically no correspondence between most of the environmental factors we studied and either shop success on the one hand or scores in mechanical prognosis tests on the other hand. Certain tendencies in the correlations indicate that an environment in which the boy does many mechanical things is better for the development of mechanical ability, so far as success in shop work is concerned, than an environment in which the father does many things. This last kind of environment is more conducive to the boy's securing a larger amount of mechanical *information* in shop courses. The correlations which substantiate these statements are:

Father's Mechanical Operations and the Manipulative Criterion	.09
Father's Mechanical Operations and the Information Criterion	.24
Tools Owned by Father and the Manipulative Criterion	.01
Tools Owned by Father and the Information Criterion	.11

A high cultural (economic) status of the family is not conducive to the ownership of tools by either father or son, or to the doing of many mechanical operations by either. The mechanical 'opportunities' of boys tend, then, to be restricted by either a large family income or by unwise budgeting of family expenditures.

Summary of Results

1. Environments of boys in a junior high school are vastly dissimilar in regard to cultural, literary, and mechanical opportunities, attitudes toward mechanical work, and family wealth. The phrase "equality of opportunity" is, generally speaking, a meaningless term.

2. The intercorrelations of seven aspects of the environment were low. Intercorrelations within the three non-mechanical aspects on the one hand and within the four mechanical aspects on the other hand, were higher than the intercorrelations between these two different sections.

3. In general, possession of economic wealth does not tend to increase the amount of a boy's mechanical environment.

4. The correlations between the seven aspects of environment both with measures of shop success and with scores in mechanical tests were low.

5. An environment in which the father does many mechanical things was more related to the information that the boy secures in the shop courses than to his manipulative skill.

6. The method of procedure followed in the research did not permit decisions concerning causal relationships. We do not know whether the doing of mechanical things about the house makes a boy more mechanical or whether a boy with mechanical ability will do more things about the house.

7. If it is safe to assume that what we did *not* measure by the questionnaire is inherited, then our data lead to the inference that mechanical ability is in large measure an inherited ability. A conclusive investigation of the genesis of mechanical ability is needed—one that shall take full account of the factors and the changes in factors covered by our questionnaire, also the amount of mechanical practice of the pupils as well as their attitudes towards mechanical work.

PREFATORY NOTE TO CHAPTER XI

This chapter deals with factors conditioning a special trait, handedness. Owing to the small number of cases thus far studied, Miss Downey's results are, of course, to be taken as suggestive, rather than conclusive.

The inference is drawn that dextrality types are probably inborn, since they are very definite in form, appear with different frequencies in the two sexes, and are manifest very early in life. Forty-nine children of age 2-9 to 5-10 tested for this investigation showed exactly the same dextrality types that adults do, although "a slight percentage of unstable reactions was found in the case of the children."

Evidence is also presented which shows that the mental age at which a child can distinguish between left and right (Binet six-year level test) may possibly be dependent in part "upon the dextrality type and the harmony or lack of harmony between hand and eye."

A third type of data deals with the coaching of nine children between mental ages of 4-10 and 5-10 who had failed the left-right test of the Binet. Four of these learned the distinction perfectly, and three of the remaining five approached "the borderline of complete success."

CHAPTER XI

DEXTRALITY TYPES AND THE PRESCHOOL CHILD

JUNE E. DOWNEY
University of Wyoming, Laramie, Wyoming

TYPES OF DEXTRALITY

Investigation of the handedness of adults carried on under financial aid from the Committee on Scientific Problems of Human Migration of the National Research Counsel, has revealed the extraordinary fruitfulness of a classification of individuals according to the types of dextrality suggested by Rife in the *Psychological Review,* 1922. Bimanual activities, such as batting and sweeping, as well as unimanual ones, such as throwing, are utilized in classification. In bimanual activities, the hand closer to the functioning end of the implement used is considered the preferred hand. One is able to establish an individual's handedness formula by noting his preferred hand in such unimanual activities as throwing, writing, eating, and the like, and in such bimanual activities as batting and sweeping. The six main dextrality types are the following: RRR, in which the right hand is the preferred hand in both unimanual and bimanual activities; RLL, in which the right hand is the preferred hand for unimanual operations but the left hand for bimanual ones; RRL, a mixed type, in which the right hand is preferred in unimanual activities but there is a mixed preference in bimanual operations, in that the right hand is dominant in bimanual activities of a high grade (batting) and the left hand in bimanual operations low in the scale (sweeping). The RRR individual throws, bats, and sweeps right-handed. The RLL individual throws right-handed, bats and sweeps left-handed. The RRL individual throws and bats right-handed, but sweeps left-handed. LLL, LRR, and LLR are left-handed types corresponding to the right-handed types just described.

It is possible, also, to describe many subclasses in a way that space here does not permit. Reference, however, must be made to the fixating eye, which is of primary importance. Handedness formulas, taken in connection with the fixating eye, give groups

which are apparently characterized by psychic traits of considerable significance. Results obtained from classification of some 1500 adults, with tentative conclusions that open up a wide field of work, have appeared in the *American Journal of Psychology*, 38: 1927, 317-367.

Inheritance of Dextrality Type

The question whether dextrality types point to native differences or are the product of the social environment cannot be answered at present, although the following facts suggest the likelihood of these types having a native, rather than a cultural, origin:

(1) The exact correspondence existing between RRR, RLL, RRL, and LLL, LRR and LLR types;

(2) The failure to find an RLR or LRL type, although a few individuals do give this combination;

(3) The existence of sex differences in the frequency with which certain types occur and also differences in the relative frequency with which certain types occur in superior and inferior groups;

(4) The possibility that structural differences (bone-measurements, and degree of finger-print symmetry for corresponding fingers of the two hands) are correlated with dextrality types.

Dextrality Types in the Preschool Child

In the endeavor to carry our range of information one step farther, forty-nine children of preschool age (2-9 to 5-10) were put through the tests previously tried on adults.[1] Exactly the same dextrality types were found for these children as for adults, although it is true a slightly larger percentage of unstable reactions was found in the case of the children. "Unstable reactions" means that variations occur in the positions of the hands in batting or sweeping or that the unimanual preference is only a mild one. It is possible, of course, that even the young child is influenced by imitation in his use of the broom or bat, or that the reaction is conditioned by pure chance. But since these young chil-

[1] Much of the experimental work for this study was carried on by Miss Caroline Wahlquist. We owe the opportunity to test the children to the kindness of Mrs. Einna Hunton, Principal of Ivinson Hall, Miss Anne Winne, teacher of the kindergarten class, and Miss Ruth Adsit, Supervisor of the University Grade School.

dren are obviously not practiced in such bimanual activities as batting and sweeping (for many of them, these performances proved entirely novel), it is difficult to believe that such high stabilization as is found in many children is due to anything other than native organization. Little girls, for example, who have previously never handled a broom may be able to manipulate it easily only with one relative position of the two hands—with the right below the left or the reverse. To attribute the extraordinary awkwardness shown in the other relative position to imitation of an absent mother or to a chance conditioning of some sort, is not convincing. In cases of mild bimanual preference, imitation is no doubt capable of establishing a type of reaction.

The relative frequency with which each handedness formula occurs in the case of children cannot be determined from the small group with whom we worked, but the figures are given since they are of interest in themselves and are, perhaps, indicative of the results that might be expected from a larger group—namely, that the higher percentage of RRL formulas for women than for men would be duplicated in the results for boys and girls.

Handedness Formula	Number of Boys	Number of Girls
RRR and LLL	7	6
RLL and LRR	6	6
RRL and LLR	6	12
Unstable	3	3
Total	22	27

Left-Eyedness

Approximately twenty to thirty percent of adults are left-eyed in sighting. The percentage is slightly larger in women than in men. It is a difficult matter to determine the sighting eye of a child. In the present investigation, the young subjects were asked to sight a colored disk through a small hole in a large sheet of white cardboard. The eyes were covered alternately and the child asked whether or not he could see the disk. In addition, we used unilateral eye-closure and eye-opening.[2] Some children give abso-

[2] Rider, W. "Unilateral winking as a test of comparative visual acuity." *Transactions of Amer. Opthal. Society*, Vol. 8, 1897-99).

lutely ambiguous results in the latter procedure; for example, they close the left eye easily and the right eye with difficulty or not at all, or they open the right easily, but not the left. If these results are in harmony with the sighting test, there is little chance of error in determining the fixating eye.

Below I give a tabular summary of the determination of the sighting eye in the case of sixty preschool children, supplementing my records by fifteen placed at my disposal by Mrs. Woolley and Miss Stutsman of the Merrill-Palmer School.

	Number of Boys	Number of Girls
Right-eyed	14	21
Left-eyed	10 (2 left-handed)	10
Inconsistent	1	4
Total	25	35

The outstanding result here is the large number of left-eyed boys.

Factors Conditioning Failure in Binet Test VI-1

The main purpose of this report is to bring the observations concerning the handedness type and the sighting eye of the preschool child into connection with results for Test 1 in the Stanford Binet six-year level—the test of the child's ability to discriminate right from left by showing the examiner his right hand, left ear, and right eye. Three correct responses are necessary if the child is to score on the test, or five of six responses if the test is expanded to include showing left hand, right ear, and left eye.

Accordingly, the Stanford Revision of the Binet Scale was given to 54 children ranging in chronological age from 2-9 to 5-10. Only two of the children who tested over the mental age of six were unable to pass this test. Both of these children were left-eyed children (one possibly fixated with either eye). This result is of interest in connection with the fact that right-handed adults who are left-eyed give evidence of a greater amount of right-left confusion than do those who are right-handed and right-eyed (or left-handed and left-eyed).

Fourteen children, ranging in mental age from 4-10 to 5-10, passed Test VI-1; eleven children, falling within the same range of mental ages, failed to pass VI-1. It is interesting to compare the handedness and eyedness of these two groups.

Formula	Passing VI-1	Failing VI-1
RRR and LLL	4	1*
RLL and LRR	3	3
RRL and LLR	5	6
Unstable	2	1
Total	14	11

*A speech case. Possibly mild unimanual preference.

	Passing VI-1	Failing VI-1
Right-eyed	8	3
Left-eyed	2 (1 left-handed)	3
? -eyed	2	5
Not-determined	2	0
	14	11

The groups are much too small to permit the drawing of any conclusion, but so far as they go, they suggest that degree of unidextrality influences the results in this Binet test.

A determination, for large groups of children, of dextrality type and of eyedness, in connection with success or failure in handling VI-1, would possibly give considerable information concerning the conditioning of success or failure in this test by special traits and would throw light upon the possible existence of at least one special ability or disability. It must, however, be borne in mind that an intelligent child learns to handle the right-left distinction by indirect means and, if unable to master it in terms of the greater kinesthetic readiness of one hand or the other, has recourse to visual or verbal tags. Furthermore, I believe that dextrality types are bound up with certain differences in intelligence and also with sex differences. For these reasons the problem of the significance of dextrality types is a very complicated one, but one which may be recommended to directors of nursery schools as deserving their fullest consideration.

The Effect of Coaching

Besides tabulating the dextrality types of the children between the mental ages of 4-10 and 5-10 who succeeded or failed in passing VI-1, we coached nine of those who failed in the test in order to discover whether or not it would be possible to teach them to tell right from left.

Four of the children (mental ages 4-4; 5-0; 5-4; 5-10) learned the distinction so that they were able to pass VI-1 without an error from six to ten different times (Nov. 17 to Dec. 10). The test was always given with six items.

The results with the other five children were not so uniform. There were days when they passed the test, and other days when they failed. Three of the children were just on the borderline of complete success, that is, they never missed more than two items out of six. One of the two remaining children was an unstable child, LLL or LLR in type, and a speech case. It did not seem possible to teach him the distinction. The youngest child (C.A. 2-9; M.A. 3-3) learned the distinction, then lost it during a period of illness, and had to relearn it. He passed the test easily at a mental age of 4-8. His verbal cues are interesting. He has been heard referring to his left leg as his "wrong" leg and to his right hand as his "easy" hand.

The results from coaching have again only suggestive value and need to be repeated on large groups of children. In general, the right-handed children who were right-eyed had the advantage.

Conclusion

In conclusion, I may say that I believe extensive testing will demonstrate that the age at which VI-1 may be passed is in part dependent upon the dextrality type and the harmony or lack of harmony between hand and eye, and that the preferred handedness type and the sighting eye are, usually, native traits.

PREFATORY NOTE TO CHAPTER XII

This chapter is the third in the series dealing with the factors that lie behind the appearance of special mental traits. The particular trait considered by May and Hartshorne, the tendency toward deceptive behavior, is of interest because it is one of the few traits of a non-intellectual sort that have been studied by experimental and statistical methods.

The material in Chapter XII represents only a portion of a larger investigation—the portion that throws some light upon the relative influence of nature and nurture in causing children to exhibit deceptive behavior. The chief issues are: do siblings resemble one another more in dishonesty than unrelated children resemble one another; if so, is the resemblance due to living in the same home, imitating the same adults, being trained by the same parents, etc., or is it also necessary, in order to explain the resemblance discovered, to assume that some portion of it must be due to inherited factors that underlie conduct?

Comparing siblings reared in their own homes with orphanage siblings by means of a 'battery' of dishonesty tests, the authors seek to separate different environmental factors. The results show that siblings do exhibit resemblance in deceptive behavior, and that this cannot be adequately explained by the social status of the homes in which they are reared. After cancelling the environmental factors, there seems to be left, in the case of deception, a degree of sibling similarity as great as that discovered by the best intelligence tests; that is, it seems as likely that deception is biologically conditioned as it does that intelligence is biologically conditioned.—*Editor*.

CHAPTER XII

SIBLING RESEMBLANCE IN DECEPTION[1]

MARK A. MAY
Yale University, New Haven, Connecticut
AND
HUGH HARTSHORNE
Institute of Educational Research, Teachers College,
Columbia University, New York City

In the course of an extensive study of deceptive behavior, the writers have been attempting to uncover possible causes of this widespread social maladjustment, and to determine their relative importance. Among such potent factors is, of course, parental influence, as mediated by the conduct of the father or mother, by their treatment of the children, by the children's treatment of one another, and possibly by biological heredity. The amount of these combined factors is roughly indicated by the correlation of siblings, inasmuch as individuals paired at random from the community would correlate zero, and family background and experience is all that siblings, by and large, have in common. But such a coefficient does not in itself discriminate between social and biological factors, when it refers to siblings living in their own homes, since both social and biological factors are at work on each pair, or each group, of sibs in ways peculiar to each family. Completely to disentangle these two lines of influence would require the measurement of brothers and sisters who had been absented from the influence of their own homes from earliest infancy. Indeed, the removal of one member of each pair from family influence soon after birth would be sufficient. Supplementary data could be secured from a study of foster children brought up in the same home and of "identical" twins (of like sex) as compared with non-identical twins and ordinary siblings.

The data reported in this article are not so comprehensive as this, and do not afford conclusive evidence regarding the relative

[1] The study here reported is part of an investigation which is being carried on by the Character Education Inquiry, Teachers College, Columbia University, in coöperation with the Institute of Social and Religious Research, New York, and will appear in more complete form in its final report.

significance of heredity and environment. But enough light on the problem is offered by the resemblances actually found to justify us in reporting them.

The Nature of the Data

The deceptiveness, or dishonesty, of several thousand school children has been measured objectively, and for the most part in quantitative terms, by placing them in real life situations and recording their responses. The situations include (a) classroom tests in which the child deceives concerning his score, (b) athletic contests in which he deceives concerning his ability to score in an 'event,' (c) party games in which he deceives concerning his achievement, (d) a home test in which he deceives concerning work done. Each of these general types of situation has one or more corresponding techniques.

For convenience of administration, the test situations are combined into batteries according to the techniques employed. There are four such batteries for measuring deception in classroom situations, one for athletic contests, one for party games, and one for home work. This study is confined to classroom and home dishonesty, since the data for siblings on other situations are at present inadequate.

The techniques used may be classified and described in terms of the four respective behaviors measured, as follows:

Behavior A. In one type of classroom situation the pupils deceive by adding more scores to a speed test after time is called. The amount which any score is thus illegitimately increased beyond what it would otherwise be is taken as the measure of deceit. There are six of these tests, affording six opportunities to cheat. The total amount of dishonest increase is reckoned as the total score in this battery.

Behavior C. In this type of situation the pupil deceives by copying from a key or answer sheet and thereby increases his score on some school test beyond what he could achieve without such assistance. Three intelligence tests are used in this battery and the total dishonest increase is taken as the total deception score.

Behavior P. This battery consists of three 'peeping' tests with techniques adapted from Voelker and Cady. The pupil deceives

by opening his eyes, or peeping, when the directions call for doing the test with eyes closed. The total amount done over and above what could be done with eyes open is the total score.[2]

Behavior H. This consists in securing help on a test taken at home, after being definitely instructed not to get any help from any source. Deceit consists in getting help. The amount is determined as in the case of Behavior C.

All these tests were standardized by giving them to a large number of children under such strict supervision that cheating was impossible or extremely unlikely. The means of such distributions were used as points of reference, and their standard deviations as units of measurement, for the cheating tests. Thus, all cheating scores are in terms of the standard deviation of honest performance. All score figures in Table I are in these terms and hence comparable for all tests.

Sources of the Data

The populations from which the sibling data in this chapter are drawn are as follows:

Population I. A suburban school system near New York consisting of five elementary schools and one secondary school in which all children in Grades IV to VIII were tested with Behavior A technique; and all in Grades V to VIII with Behavior C and Behavior H techniques. In this town both members of each pair of siblings either attended the same elementary school, or, more frequently, one of the pair attended an elementary school and the other the secondary school. This community offered a wide range of home environments, ranging from homes of comfort and affluence to those of abject poverty.

Population II. A much smaller suburb of another city in which there were two schools. All children in both schools in Grades IV to VII were tested with Behaviors A and P techniques. In every case both members of each pair of siblings attended the same school. The range of home environments here was not so wide, lacking the lower extreme.

[2] The total amount done is used in this study, as the amount done honestly was too small to affect the correlations.

Population III. Four public schools in a congested city district were tested with Behaviors A, P, and H in Grades IV to VIII, except that Behavior H was not given to fourth-grade children. These children were racially homogeneous, with home environment rather uniformly low in the economic and social scales. Both members of each pair of siblings attended the same school.

Population IV. In a mid-western town of 200,000 inhabitants Grades V and VI of two representative elementary schools and Grades VII, VIII and IX of a junior high school were tested with Behaviors C and H. These children came from a rather wide range of homes.

Population V. This is a junior high school in a large suburb. All pupils in Grades VII, VIII and IX were tested with Behavior P. The home backgrounds of these children were distinctly above the average.

Populations VI and VII. Two groups of orphans were tested. Both were in orphan homes and both represented the same type of environment as found in Population III. One institution, VI, is in a large city; the other, VII, is in the country. In the former we tested one hundred pairs of siblings with Behavior C; in the latter we tested forty-five pairs with Behaviors A and P.

RESULTS

Table I conveniently assembles the pertinent data and statistical results. Column 1 gives the population, referring by Roman numeral to the groups already described. Column 2 is the number of pairs of siblings in each case. Column 3 gives the correlations between pairs.[3] Column 4 gives the standard deviations of the scores of all the cases. Column 5 gives the standard deviations of the differences between the paired scores. Column 6 gives what the

---

[3] In computing these correlations all combinations were inserted where there were more than two siblings in a family. Also each pair was entered twice; each individual thus appears on both the X and Y axis of the scattergram, making the plots symmetrical, with N equal to the number of entries. The work of correlating was facilitated by the use of the formula: $r = 1 - \frac{\sigma^2_d}{2\sigma^2}$ which is derived from the Pearson product-moment formula, σ_d (Column 5) being the standard deviation of the differences and σ (Column 4) the standard deviation of the entries.

sigmas of the scores would be if the pairs were selected at random and the resulting correlations were zero. Column 7 is Column 6 minus Column 5. Column 8 is Column 7 divided by Column 6.

TABLE I.—STATISTICAL RESULTS OF SIBLING STUDY*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Population	N (Prs.)	r	σ	σ d	σ d o	6–5	Pct. of 7 to 6
Behavior A Speed	I	76	.220 ± .05	3.27	4.09	4.62	.53	11
	II	70	.292 ± .05	1.59	1.89	2.25	.36	16
	III	224	.208 ± .03	3.42	4.29	4.83	.54	11
	I, II & III	370	.225 ± .02	3.00	3.72	4.24	.52	12
	VII	43	.300 ± .07	2.65	3.12	3.75	.63	16
Behavior C Copying	I	108	.445 ± .04	1.46	1.54	2.02	.53	26
	IV	138	.433 ± .03	1.50	1.59	2.12	.53	25
	I and IV	246	.440 ± .03	1.38	1.46	1.96	.50	25
	VI	94	.333 ± .04	1.82	2.09	2.57	.48	18
	+½	38	.333 ± .07	1.73	2.00	2.45	.45	18
	−½	46	.333 ± .06	2.03	2.34	2.86	.52	18
Behavior P Peeping	III	239	.322 ± .03	2.62	3.04	3.70	.66	17
	II	74	.271 ± .05	3.02	3.65	4.27	.62	14
	V	89	.271 ± .05	2.80	3.39	3.96	.57	14
	II, III, V	402	.400 ± .02	2.86	3.15	4.04	.89	22
	VII	43	.300 ± .07	2.73	3.22	3.86	.63	16
Behavior H Help	I	104	.641 ± .03	2.47	2.07	3.49	1.42	40
	IV	88	.472 ± .04	2.25	2.30	3.18	.88	27
	III_1	94	.695 ± .03	2.13	1.69	3.01	1.32	43
	III_2	55	.832 ± .02	2.73	1.58	3.86	2.28	59
	I, III, IV	345	.705 ± .01	2.62	2.01	3.70	1.70	46

*Note the following relations between the columns of Table I:

Column 3: $r = 1 - \frac{\sigma^2_d}{2\sigma^2}$

Column 4: $\sigma = \frac{\sigma d}{\sqrt{2(1-r)}}$

Column 5: $\sigma_d = \sigma\sqrt{2(1-r)}$

Column 6: $\sigma_{do} = \sigma\sqrt{2}$

Column 8: $\frac{\text{Col. 7}}{\text{Col. 6}} = 1 - \sqrt{1-r}$

INTERPRETATION

Before indicating the bearing of the facts of the table on various hypothetical explanations of sibling resemblance, certain features should be noted.[4] The important comparisons are between the dif-

[4] Column 8 shows the percentage of restriction in range of difference between siblings compared to range of differences as found in uncorrelated pairs. The greater the figure in Column 8, the greater the restriction of range in differences or the greater the resemblance of the pairs. The figures of Column 5 depend on both the correlation between pairs and the sigma of the scores. These sigmas are in the same units in each section of the table and are comparable among the tables. We may thus compare the resemblances shown by Columns 7 and 8 with the resemblances indicated by the correlations alone.

ferent behaviors on the one hand and the different populations on the other. First, as to behaviors, Behavior H, which is the test taken at home, elicits the greatest resemblance between siblings, doubtless partly because of collusion. Behavior C comes next, perhaps because the copying type of deception is more consciously dealt with by parents. Behaviors A and P come last, but still show genuine resemblance.

Second, as to populations, the orphans, VI and VII, were not given Behavior H, but of the other populations, the congested population, III_1 and III_2, shows the greatest resemblance. In III_1 and III_2 two schools are combined, a better and a worse in each case. The average r for all of III is .763. In the other behaviors little difference among the normal populations is found, but in the case of Behavior C, there is considerable difference between the normal and the orphan groups.

Possible Explanations of Sibling Resemblance

We have now to account for these likenesses and differences among the populations and point out their significance for the interpretation of the sibling resemblance. We shall discuss three possible explanations: collusion, common environment, and heredity.

1. *Collusion:* This is possible only in Behavior H, the task done at home. Here we may expect that some children will either help one another or encourage each other in the use of forbidden aids. The size of the sibling r's bears out this expectation. For all populations it is .705 and runs as high as .83 in one group containing fifty-five pairs. This is considerably higher than the reliability of this single test and indicates that there is greater likelihood that two siblings will cheat in equivalent amounts on a single occasion than that the same child will cheat in the same way on two occasions. If mere general home influence were the prepotent influence, this would not be the case. As a matter of fact, general social-cultural level, as determined by an experienced case worker, correlates .03 with Behavior H, showing that whatever makes siblings resemble one another in deception on home work is not the kind of thing in which homes as such differ from one another, so far as we have measured them.

Furthermore, if siblings helped each other on the home test or got help from a common source, one might expect the older to exert the greater influence. This cannot be shown from our data, but it is of interest that, although there is no great difference between the mean of the older and the mean of the younger, yet in a population of 119 pairs, 9 percent were identical, in 61.5 percent of the pairs the younger cheated more than the older, and in 29.5 percent of the pairs the older cheated more than the younger. The older and younger were nearly alike in variability.

2. *Common Environment.* There are, of course, other home factors than collusion, direct assistance or example that make for honest, rather than dishonest, behavior, or vice versa (such as code, ideal, or general social stability), so that one might expect that two children reared in the same home would, by sharing certain common experiences, become more alike than if one were reared in one environment and the other in another. Of course, 'the home' is not really identical, even for its own children, but certainly the range of social experience offered by a large number of homes is far wider than that offered by one. If deception is correlated with these variable factors, the inevitable result will be that siblings will resemble one another more than the children of other families by an amount expressed by the coefficient of correlation between members of pairs.

The populations reported in the table present about as wide divergence as is likely to be found in cities and suburbs. The inclusion of rural groups would extend the range, but it is wide enough as it is to produce correlations so far as these may be due to social influence, and to allow for variation in the size of correlations as longer or shorter sections of the entire range are taken for comparison. So far as deceptive behavior is built into habit systems and is not a mere matter of collusion or example in the home, it will appear in the correlation of tests taken away from home. We have already pointed out the existence of such correlations as are found in Column 3 of our table.

Let it be clear that if the correlation between home environment and deception on tests taken away from home is due to influences of the home which vary with social levels, then,

1. The wider the range of homes reported in a given population, the greater will be the variability in deception; and the narrower the range of homes, the smaller the variability in deception.

2. In like manner, the range of homes in the different populations will show a correlation with the different sibling correlations; that is, the low sibling r's will be observed in the populations with the narrower range of homes; and the larger sibling r's will be found in the populations with wider range of homes.

Altering the Range of Environment

There are three available ways of expanding or contracting the range of environment in order to test this second hypothesis of the social causation of our r's. One is to select a population which all judges would agree is relatively homogeneous. A second method is to measure home background and by applying the partial correlation technique keep this influence constant. A similar method is to select as a basis of study the cases falling within certain sections of the home background scale. All three methods have been employed and will be reported.

Method 1: First it is necessary to indicate roughly the range of socio-economic background of Populations I to VII.

The ranges of home background reported in the various populations tested have been determined partly by test[5] and partly by observation and common consent. Their relative ranges are shown in Figure I.

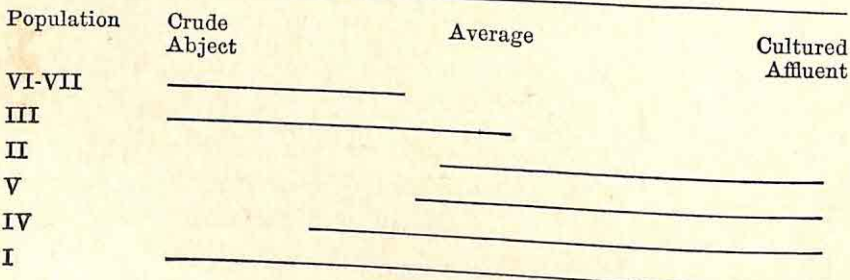


FIG. I.—RANGE OF HOME BACKGROUND FOR THE SEVEN POPULATIONS

[5] The test used was a questionnaire standardized and validated by Dr. V. M. Sims (and issued by the Public School Publishing Co.), as a measure of socio-economic level of homes. It is administered in school. Sims reports its reliability as .94.

The range of background for the two orphan groups is really much narrower than indicated in the chart, for no matter what the range of background from which they came, they had for some time been living in *one* environment, as though all were children in one big family.

In width of social range the populations rank roughly: VI-VII, III, II, V, IV, I. To the extent that the social factors thus distributed are determinative of deception, we should expect a corresponding variability in deception range, or deception sigmas, and in deception r's. The table shows that this expectation is not realized. For convenience the sigma and correlation ranks are gathered in Table II, the columns of which are as follows: Column 1 gives the rank order (slightly modified) of the various populations in spread of home background as in Fig. I. Columns 2, 4, 6, and 8 give the (modified) rank order of these same populations in the size of sigmas of their deception scores for Behaviors A, C, P, and H, respectively. Columns 3, 5, 7, and 9 give the rank-order in size of r's between siblings for the same four behaviors.

TABLE II.—RANK ORDER OF POPULATIONS IN SOCIAL VARIABILITY, DECEPTION VARIABILITY, AND SIBLING CORRELATION

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Range	σ_A	r_A	σ_C	r_C	σ_P	r_P	σ_H	r_H
VII.........	1	3	2	......	......	1	2	......	......
VI.........	1	......	......	2	1	......	......	......	......
III.........	2	6	1	......	......	1	3	2	5
II.........	3	1	2	......	......	3	1	......	......
V.........	4	......	......	......	......	2	1	......	......
IV.........	6	......	......	1	3	......	......	1	1
I.........	8	5	1	1	3	......	......	2	3

From Table II it can be seen that there is no correspondence between deception range and social background range, except in the case of Behaviors P and H, which show a slight tendency toward equivalence in range. We infer, therefore, that there are factors other than general socio-economic level at work to promote deceptive habits.

The same conclusion is reached from the relations among the r's, for in only two cases do the r's increase with range of background, *viz.*, Behavior C and Behavior H. In the latter case Population

III is an exception to this tendency, showing the highest r between siblings in a narrow range of home background, though here the difference may be accounted for by collusion.

Even the orphan population shows a greater variability than Population II and a larger sibling r than Populations III or I in Behavior A, a greater variability than Populations IV or I in Behavior C, and a larger sibling r than Populations V or IV in the case of Behavior P. Also in the orphan population, VI, the children who had been living in the institution more than half their lives ("+½" of Table I) show the same sibling r as those who had been in the institution less than half their lives ("−½" of Table I), and only slightly less variability. If gross variation in environment is producing the sibling correlation, one would expect the r to be lower in the case of those who had been longer in the institution.

Method 2: The second method of testing the environmental hypothesis involves securing some measure of the variability of home background with a view to determining its correlation with deception. If we can arrive at even an approximation of the correlation between variation in homes and deception, we can then use the partial correlation technique and hold the home variation constant in the sibling r's.

We have made use of three ways of measuring the socio-economic level of homes. As a first crude approximation we used the occupation of parents. We ranked the occupations in three groups—professions, trades and crafts, and unskilled labor. The relation between this grouping and the deception (Behavior C) of the children is about .30 (coefficient of contingency). This determination was made on the two populations, I and IV, where the range of homes is widest.

In two populations, II and III above, we measured the socio-economic levels of the pupils with the Sims score card. In one of the lower ranges of homes, Population III, it yields a correlation with cheating of only .115. In the upper range the r is about .20. If all ranges are combined so that we have a kind of unselected sample of homes, the correlation between Sims score and cheating rises to .30, which is about the same as we got with the crude occupational measure.

In another study we employed a trained social worker to visit the homes of the most deceptive and the least deceptive. She made

a very careful study of over 150 homes to find, if possible, the details of experience and environment that were influencing the conduct of the children. These homes were grouped in fifteen levels according to their quality as homes. The correlation between these homes, thus rated, and the deception of the children is .30,[6] again the same as found by the other two procedures. But the regressions here are not linear. The correlation ratio of homes to cheating shows an eta of .454, and that of cheating to homes, one of .440. In the case of the Sims scores the regressions were linear. We shall return to this point later.

These figures indicate that the correlation between range of homes as measured in these three ways and cheating in school is around .30. We can now find the partial r's between siblings in deception with home background (as far as we have measured it) constant. But since the above r of .30 represents a wide range, we should use only r's from Table I which are also on wide ranges. The resulting partials are given in Table III.

TABLE III.—SIBLING RESEMBLANCE WITH HOME BACKGROUND CONSTANT

Behavior	Population	Observed r	Partial r*
A.	I, II, III	.225	.143
C.	I, IV	.440	.385
P.	II, III, V	.400	.341
A+C+P.		.470	.394

*Based on the assumption that the r between siblings in home background is 1.00. See footnote No. 7.

The r's indicate that children who come from homes of the same general social level still show correlations between siblings.

Method 3: As a check on this we took from Populations II and III all those who deviate more than about two sigma from the Sims socio-economic means. We found the r between the siblings for classroom deception beyond this extreme in the upper range to be .35 and beyond this extreme in the lower range .30. We then took a 'slice' out of the middle, including all those between about

[6] We recognize that a correlation obtained from the measures of the extremes in one variable is somewhat spurious. But our aim here is to arrive at only an approximation of the correlation between variation in homes and deception.

+.3 sigma and −.3 sigma and still got a correlation for classroom deception of about .35. These determinations check and verify the partials.

Further Criticism of the Common Environment Hypothesis

Our study of the influence of a common environment on sibling resemblance has so far led us to the conclusion that, while such factors as we have been able to measure or observe are doubtless significant, they do not account for all the resemblance found—that there is left about as much likeness as is found in the case of sibling resemblance in intelligence (as will be shown presently). Before turning to the consideration of the other proposed explanation, there is one more technique for the evaluation of our data which will be suggestive in showing how inadequate the observed relations between deception and home background are to account for sibling resemblance.

In the first place, it can be shown that the correlation between home background and deception must equal the square root of the correlation between siblings before it can completely account for it. The relation is $r_{sc} = r^2_{ch}$, in which r_{sc} is the correlation between sibs in cheating and r_{ch} the correlation between home background and cheating.[7]

We have estimated that the correlation between siblings for Behaviors A, C, and P, combined, is not less than .47. If this is to be wholly accounted for by virtue of the type of social influence measured, then the correlation between home background and cheat-

---

[7] This simple relation is derived as follows:

Let $r_{sc} = 1 - \frac{\sigma^2 dsc}{2\sigma^2 c}$, where σ^2dsc is the standard deviation, squared, of the difference between siblings in cheating.

Let $r^2_{ch} = 1 - \frac{\sigma^2 ac}{\sigma^2 c}$, where σ^2ac is the standard deviation, squared, of the arrays of cheating for any home level, and σ^2c the standard deviation, squared, of cheating.

It can be shown that $\sigma^2 dsc = 2\sigma^2 ac$.

Then, by substitution, $r_{sc} = r^2_{ch}$.

This can be checked by the formula for partial correlation in the case of siblings, which is $r_{sc.h} = \frac{r_{sc} - r^2_{ch}}{1 - r^2_{ch}}$. This formula is a special case of the usual formula for partials and is applicable on the assumption that the correlation between siblings in environment is 1.00.

Now, when $r_{sc} = r^2_{ch}$, then $r_{sc.h} = 0$.

ing should be $\sqrt{.47}$, or .685. Our measures and our estimates by an expert afford a correlation of only .30 between deception and home background, which is enough to account for an r of only about .10 between siblings. In other words, there is considerable resemblance still to be explained.

The second application of this technique leads up to our final suggestion.

As has been noted, our measures of home background are not perfect. It may well be that, with more refined methods for revealing other factors than those included in the three methods described, the correlation between deception and home background might be increased to a point sufficient to account for the correlations between siblings. Suppose for a moment that this has been done. Then in the case of Behavior P, where $r_{sc} = .40$, the correlation between cheating and home background would have to equal $\sqrt{.40}$ or .63 to account fully for the .40. Let us suppose that this is the case, and that the r of .63 is based on an unselected population covering a wide range of environmental levels. Then, if we reduce the variability of the home background (as measured by our hypothetically adequate test) by one half, we shall automatically reduce the r of .63 to .45, and the sibling r from $.63^2$, or .40, to $.45^2$ or .21.[8] This .21 is the predicted r between siblings from a population when the standard deviation of the environment variable is reduced one half and the r between cheating and the full range of the environment is .63.

Now, as a matter of fact, taking our rough range-of-background chart as our measure of the relative range of different groups, Population III shows about half the variation of the total range, and Table I shows that the correlation between siblings for Behavior P in Population III is .32, which is only .11 higher than the predicted .21. Also, Populations II and V, likewise greatly restricted in scope, show r's in each case of only .27. These three r's have P.E.'s of .03 for III and .05 for II and V, so that the predicted r of .21 is within 4 P.E. of one of the observed r's and within 2 P.E. of the other two.

The meaning of this is that, in the case of Behavior P (peeping), whatever resemblance there is between siblings may well be ac-

[8] For formula, see Kelley, T. L., *Statistical Methods*, p. 225, Formula 187.

counted for by such environmental factors as are included in our measures.

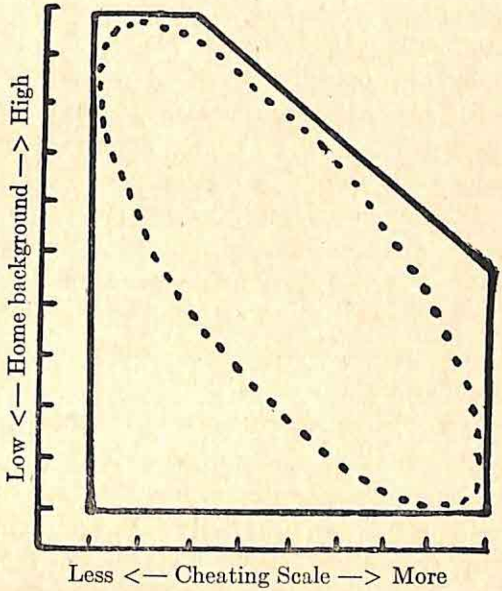
But how about Behaviors A and C? In Behavior A (speed), restricting the environment does not reduce the r's in the way called for by our formula. Whatever resemblance siblings show has little, if anything, to do with the gross environmental factors.

Behavior C (copying from answer sheet) shows a slight fall in sibling r when range of background is restricted. The r_{sc} here is .44. To be wholly accounted for by home variation, the r_{ch} must be .663. Reducing the variation by three quarters, as for Population VI-VII, drops this theoretical r_{ch} to .30 and the corresponding r_{sc} to $.30^2$, or .09. The range of background in Population VI-VII (orphans) is surely not more than a quarter of the entire range as to gross factors, yet the actual r_{sc} is .33. Evidently, there is something at work in creating sibling resemblance that is not accounted for by the gross likeness in sibling environment. However, if the background is reduced in variability by only one half, the r_{sc} only drops to .25.

In this whole discussion of common environment as a possible cause of sibling resemblance we have been dealing with variations in the general socio-economic-cultural levels of homes, the kind of difference found between congested districts and comfortable suburban homes. We have shown that it may account for resemblances in Behavior P, but not wholly for resemblance in Behaviors A and C. Let us turn then to the third tentative explanation of this resemblance.

3. *Heredity*. We have mentioned the probability that homes may differ widely in other respects than socio-economic or cultural levels. There may be wide variation among homes of the same socio-economic level in respect to attitudes toward children, general stability and adjustment, or even in codes and ideals. These subtle factors may be correlated with deception to a much greater degree than those factors which are distributed in accordance with general socio-economic level. The existence of such factors, which is indicated by our sibling correlations, is forcibly illustrated by the fact that whatever makes for dishonesty is left behind as we go up in the social scale, but whatever makes for honesty is not dropt out as we go down in the social scale. On the lower social levels, there-

fore, the correlation between deception and home background is zero. If deception is plotted against home background, the shape of the scattergram is as indicated by the solid line of the accompanying diagram. If deception correlated .75 with home background ratings, the shape of the scattergram would be about as indicated by the dotted line.



Had all possible environmental factors been kept constant in the partial r's of Table III, we might claim that the remaining sibling resemblance in the matter of deception is due to biological factors. This statement of the hereditary factor is quite unsatisfactory, however, as nurture plays a rôle in the development of all measurable mental traits and is not 'eliminated' in any true sense from any group of factors. Some elements in behavior, however, are usually regarded as less modifiable than others and more directly the result of biological growth than of interaction with the environment. General intelligence, for example, although its particular modes of exhibition are socially determined, is usually thought to change very little with training. Does the tendency to deceive belong in the same category, or is it due to particular environmental relations in which the prepotent factor or factors are subject to modification as are ideas or ideals or the ability to

read or to translate Latin? The relatively innate factors[9] would include temperament, emotional stability, power of inhibition, response to persons as compared to things or ideas, and general intelligence.

It is commonly supposed that the correlation between the I.Q.'s of siblings in a homogeneous age population of unrestricted range is around .50. Miss Hildreth[10] summarizes work done up to 1925, showing that the actual results on over two thousand pairs run from .30 to .63. These yield an average of .48. Gordon's data on 219 pairs of orphans, when recalculated by Elderton by entering each pair twice in the scattergram, give a correlation of .467. Miss Hildreth reports .322 as the obtained r between 253 pairs of orphans. These two average .388.

Our deception correlations, when corrected for attenuation, show comparable results. Extracting from Table I the r's obtained on populations of widest range and correcting these for attenuation, we get Table IV.

TABLE IV.—COEFFICIENTS OF CORRELATION BETWEEN SIBLINGS IN DECEPTION, CORRECTED FOR ATTENUATION

Population	Behavior	Raw r	r Corrected for Attenuation
I, II, III	A	.225	.256
I, IV	C	.440	.510
II, III, V	P	.400	.470
Theoretical	A + C + P	.470	.495
Orphans VII	A	.300	.341
Orphans VI	C	.333	.387
Orphans VII	P	.300	.353
Theoretical	A + P + C	.370	.390

Thus, the resemblance between school siblings in deception is about the same as it is in I.Q. Furthermore, it shows the same amount of independence of home background range as is shown by resemblance in intelligence. These facts may mean that the tendency to deceive is conditioned by nature and nurture in about the same proportion as is general intelligence.

[9] It is of some importance that liars have been found to be "more suggestible, more imaginative, less steady in motor control involving an emotional disturbance." See the study in untruthfulness by Slaght, University of Iowa Studies in Character Education No. 4, referred to by Starbuck, E. D., *Religious Education,* January, 1927.

[10] Hildreth, G. H., *The Resemblance of Siblings in Intelligence and Achievement.* Teachers College, 1925.

SUMMARY

Seven hundred thirty-four pairs of siblings scattered among seven school populations were measured with four different types of deception tests. The general socio-economic level of the homes was measured in two populations by a scale made for this purpose and by the careful estimates of a trained school visitor in one population. Two populations were orphans from one of the communities already measured. The range for the other two populations was estimated.

These siblings were found to resemble one another in deception, whether the opportunity to cheat was offered at home or at school. The resemblance in deception on the test taken at home was shown to be due chiefly to collusion. Two theories were proposed to account for the resemblance between siblings in deception on tests taken in school when collusion was impossible, *viz.*, (1) the influence of environment, through which siblings develop the same standards and habits of school behavior, and (2) the presence of a common hereditary factor or group of factors in children of the same parents showing itself in similar deceptive behavior. These two theories are not mutually contradictory.

The data assembled show clearly that the coarser differences among homes in socio-economic status, although associated somewhat with differences in deception, do not fully account for likenesses between siblings. So far as the home training is responsible for the resemblances observed, it must be due to subtle factors not incorporated in the home ratings of a trained case worker. Furthermore, since orphan siblings show resemblances which cannot be attributed to present home background, what home influence, if any, is at work must be exercised in early childhood.

A biological factor is thus suggested as a possibility which is made the more plausible in view of the fact that the resemblance of siblings in deception is about the same as that between siblings in intelligence, and shows similar stability amid home backgrounds of varying range as is shown by resemblance in intelligence. Whatever arguments are adduced for or against the biological causation of sibling resemblance in intelligence are applicable also to the interpretation of sibling resemblance in deception.

PREFATORY NOTE TO CHAPTER XIII

One may not at first see the relation of this chapter to the nature-nurture problem, but the relation will be more evident if the problem is conceived as that of the possibilities and limitations of training. If training of a particular mental function has no influence on efficiency in other mental functions, it is obvious that the total effect of the educative processes, in school and out, will be less than if transfer takes place. It is not, to be sure, a question of the improvement or non-improvement of original endowment, but of the extent to which original endowment in a particular line is realized by training in other lines. Life is too short and the kinds of mental performances are too numerous to enable anyone to develop every capacity to its limit by specific training; so it is important to know whether and to what extent general efficiency is influenced by the development of special efficiency.

The conclusion of the author that some degree of transfer does take place, and that "some of the most important agencies of transfer are to be found among the higher-level relations, in generalized attitudes, words, ideals, sets, and ways of going about mental operations," seems to be justified by the experimental literature on the problem. If this is true, it is possible that formal training may have an effect upon the I.Q. We would certainly be justified in so concluding unless it could be shown that the ordinary experiences of everyday life and of schooling of any reasonable kind themselves give the total amount of general training which it is possible to secure. Whether the more formal kinds of specific training are necessary to accomplish this result must still be held an open question.

The summary has been organized in six sections. Readers who are primarily interested in the final conclusions will save time by turning directly to the sixth section, or better, to the fourth, fifth, and sixth sections. Readers who question the validity of the summaries and conclusions therein presented or who are unfamiliar with the history of this aspect of educational psychology will perhaps find it worth their while to peruse the entire chapter.

—L. M. T.

CHAPTER XIII

THE TRANSFER OF TRAINING

GUY M. WHIPPLE
Danvers, Massachusetts

It is the purpose of this chapter to set forth the facts concerning that aspect of the general problem of the possibilities and limitations of training which is characterized by the stock phrase, "the problem of formal discipline," or "the problem of the transfer of training." It will be desirable to consider (I) what the problem is, (II) the methods and materials used in its investigation, (III) the results of typical investigations, (IV) what has been proved concerning the amount of transfer, and (V) what has been proved concerning the agencies of transfer, from all of which may be drawn (VI) certain inferences concerning the pedagogical significance of transfer.

I. The Problem of Transfer

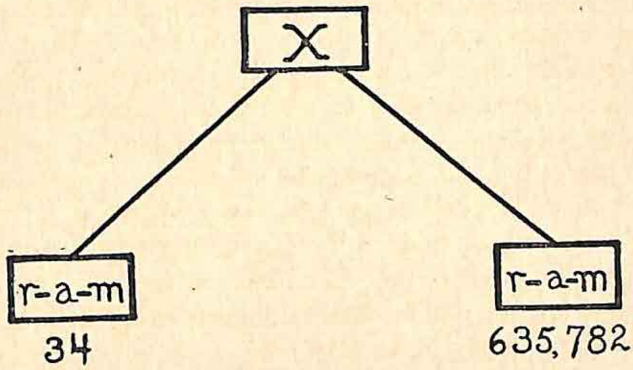
Suppose a stimulus, *r*, impinged upon some sense organ and, after passing through some central portions of the nervous system, some associative neural connections, *a*, finally, by means of some motor neurone, set into action some motor response, *m*. This schema may be thought of as picturing the theoretically simplest sort of sensory-motor response. Suppose that this sequence of activity—*r*, *a*, *m*,—at first operated haltingly or feebly or inaccurately, but that its operation resulted in no arousal of pain, rather, on the contrary, produced some sort of satisfaction to its possessor. If, now, *r* is repeated, there is a definite tendency toward the arousal of the same sequence of activities, and if *r* is repeated often enough or intensely enough or in sufficiently close succession, we may expect the sequence eventually to operate unhesitatingly, strongly, and accurately. We then say that this response (*m* following upon *r*) has been improved by practice and eventually, if it continues to operate in this improved manner, that it has been 'learned.' This process is the process of training, reduced to the

simplest terms. Thus, for example, the infant 'learns' after a number of trials to suckle his mother's breast successfully or the schoolboy learns to answer "six" to the question "How much is four plus two?"

Now, it is possible to conceive of the entire operation of the nervous system (or of our mental activities, if we prefer to talk in psychological terms) as representing an enormously complex aggregation of hundreds of thousands of such 'stimulus-association-response' acts. The suckling response might have been Number 34 that the infant encountered and perfected; the answering "six" to "four plus two" might have been Number 635,782 of the series making up his life. It is also possible, as a matter of psychological theory, to assume that each one of these hundreds of thousands of responses, or learned acts, is an independent unit, that it operates in and by itself, that it has to be experienced and perfected *ab initio,* without any reference to the other hundreds of thousands of responses. On this assumption, the particularity, or the specificity, of the individual response is complete; its relationship to other responses is zero psychologically (whatever it might be logically). On this assumption, the problem of transfer disappears, since it becomes obvious that there can be no transfer; the improvement of any single response will never bring about the slightest improvement in any other response.

It is possible, once more as a matter of psychological theory, to assume that any one of the hundreds of thousands of responses, or learned acts, may be to some extent neurologically (or psychologically) 'hooked up with,' or interrelated with, numerous other responses. On this assumption, no given response can be positively guaranteed to be purely particular in its operations, because the chance that any given response is so completely dissimilar in its nature and operation from every other one of hundreds of thousands of responses as to be thus totally unrelated to any one of them is obviously exceedingly small. This is only another way of saying that responses actually are complex and mutually overlapping. It is only another way of saying that the simple schema —*r, a, m*—with which we started is inadequate to the real situation. Thus, it requires no wild stretch of the imagination to suppose that if Response No. 635,782, previously mentioned, had been perfected,

and if Stimulus No. 635,783 were "How much is two plus four?" the appropriate responsive act—saying "six"—might be forthcoming with no previous training at all. Or, again—drawing heavily, perhaps, on our imagination this time—supposing that in perfecting Response 34 (learning to suckle effectively) the infant in some vague organic way discovered that it "paid to keep at it"; then, is it not conceivable that at the age of seven the same youngster might on that account exhibit a very creditable degree of persistence in learning his sums? Let us suppose, anyway, for the sake of argument, that that is what happened. If so, we must rearrange our original schema, at least so far as to add some factor, *X*, which stands for: "When confronted by an obstacle, keep



at it." Now, all we have to assume is that learning the sum of four and two presents a slight obstacle, which immediately brings into the learning process the by-now-well-correlated "keep at it," with the result that the new acquisition is accomplished, though conditions are such that otherwise it would not be accomplished. Here, then, we do have real transfer of training. The factor, X, common to the two seemingly utterly diverse situations (34 and 635,782), which was developed or improved when 34 was operating, is now a potent factor in making 635,782 operate. This is the sort of thing that takes place, according to many writers who champion the 'common element' theory of transfer.

Between these two theoretical extremes, the complete specificity of psychophysical activities on the one hand and the complete interlocking of them on the other, we may, of course, assume the operation of any assignable intermediate degree of interrelationship. And since no one seriously champions either of the extreme

hypotheses, the actual problem of the transfer of training becomes *largely a problem of degree*—and, of course, of adequate description of the underlying conditioning factors. That is, how much transfer is there? Under what conditions does it take place? Under what conditions does it not take place? In short, there are two chief problems confronting the experimentalist—the *amount* of transfer and the *agencies* of transfer.

Both of these problems are of decided theoretical interest—it is difficult, for example, to see how one can develop his notion of the nature of mental operations without committing himself to some sort of formulation of their organization into certain more or less independent and other more or less interrelated mental functions.

Both of the problems are likewise of decided practical interest, as, for instance, in laying out a satisfactory program of mental training, in deciding what forms of pedagogical instruction are needed to bring about specified improvements in mental efficiency, and so on. These pedagogical issues are perfectly familiar to our readers, especially as they become formulated in terms of the various school subjects, *e.g.*, "Does training in geometry improve reasoning?" "Does training in Latin assist the pupil in acquiring French or German?" "Does penmanship transfer?" meaning "Does the child continue in his after-school life to follow the technique of writing taught him laboriously in the grades?" etc.

Such, then, is the problem of the transfer of training. In the simplest terms it is: Does the improvement in a single response to stimulation entail improvement of any other response? In more complex terms it becomes: Does training in a complex and more or less related series of responses (like that involved in learning a given branch of subject matter) entail perceptible increase in capacity to learn some other series of responses (like those involved in learning some other branch of subject matter)?

II. General Methods and Materials Used in the Experimental Investigation of Transfer

The literature of experimentation in the field of transfer is voluminous; to attempt a comprehensive summary of all the investigations that have been reported would far exceed the space

allotted to this chapter. As a matter of fact, such an attempt is unnecessary because several such summaries have already been made.[1]

Our aim here is to describe certain investigations that may be regarded as typical of the methods and materials that have been used and of the interpretations that have been drawn.

First, however, a few words about the methods, the materials, and the subjects used in the experimental work.

1. Methods

With respect to methods, it is possible to distinguish (a) the individual method, (b) the one-group method, (c) the two-group method, and (d) the three-group method.

(a) The *individual method* consists, of course, of experimenting done by one person who measures his ability in some specified activity, next trains himself for some time in some more or less related activity, and then measures once more his ability in the specified activity, with the intent to discover to what extent, if any, the period of training has affected his performance in the specified activity. The early work of William James, in his study of transfer of training in the field of memory, illustrates this method. Naturally, in the hands of a skilled psychologist, this method may yield results of significance, especially when fortified by introspective comments, but methodologically it can scarcely be defended.

(b) The *one-group method* differs from the individual method primarily merely in increasing the number of subjects. An obvious illustration is afforded by the early experiments reported

[1] The reader will find useful summaries, for example, in the following:
Colvin, S. S. *The Learning Process*, Chs. XIV-XVI.
Heck, W. H. *Mental Discipline and Educational Values.*
Henderson, E. N. *Textbook in the Principles of Education*, Ch. X.
Hewins, Nellie P. *The Doctrine of Formal Discipline in the Light of Experimental Investigation*, pp. 4-48.
Rugg, H. O. *The Experimental Determination of Mental Discipline in School Studies*, Ch. I.
Starch, D. *Educational Psychology*, Chs. XIII-XIV.
Thorndike, E. L. *Educational Psychology*, Vol. II, Ch. XII.

Of these, perhaps the most comprehensive and useful is that of Rugg, who, by means of charts, has prepared an excellent synoptic summary of the essential features of all the studies prior to 1915.

by Thorndike and Woodworth in which a group of subjects practised estimating the areas of rectangles until a "very marked improvement was attained." The same group then estimated areas of the same general size but of different shape and areas of the same shape but of different size. By the use of this method individual irregularities in the effect of training are more or less eliminated or compensated, but there is no way to determine accurately how much of the improvement is due to a real transfer effect, how much is due to improvement within the test series itself.

(c) The *two-group method* has been generally employed in all the more recent experimentation to meet the objection just cited. In principle, the method is simply this: the subjects are divided, on some desired basis, such as age or general intelligence or previous training, into two equivalent groups, of which one is known as the "Trained," or the "Experimental" group, the other as the "Control" group. Both groups take both the preliminary and the final test series, but only the former takes the intermediate training series. If the division of the groups is carefully done, the difference between the performance of the trained group and the control group in the final test may be expected to indicate the amount of transfer effect, uncomplicated by possible effects of practice in the preliminary series or of maturation, incubation, and similar influences that might conceivably cause the final test results to differ from the preliminary test results.

(d) The *three-group method* is an elaboration of the two-group method which may well be justified under certain conditions of experimentation. This method follows the plan of the preceding one, but there is added a third group (equivalent in characteristics of its personnel to the others) which takes none of the tests except the final one. If, then, unavoidable changes are present in the final tests, as in the weather, in the school program, in the experimenters, in the place, etc., so that conditions differ from those prevailing in the preliminary test, their effect may be studied and allowed for. An illustration of the use of this method may be found in C. P. Wang's investigation of the general value of sense training.[2]

[2] Wang, C. P. *The General Value of Visual Sense Training in Children.* Educational Psychology Monographs, No. 15, Baltimore, 1916.

2. Materials

By "materials" I refer to the nature of the mental activities chosen as the objects of the investigation. The investigations may be conveniently assembled, for descriptive purposes, into four groups, viz., (a) those dealing with memorizing, (b) those dealing with various aspects of perceiving, including sensory discrimination and apprehension, (c) those dealing with voluntary effort and motor adjustments, (d) those dealing with schoolroom activities and attitudes. It might be possible to make a fifth group to include investigations dealing with so-called "cross-education," or the transference of practise gained in one member of the body to a symmetrical member.

3. Subjects

There has been a decided tendency to carry on the investigation of the transfer of training with adults as subjects. Thus, taking 25 of the studies reported prior to 1916, we discover that only seven of these were conducted with school children, while of the remaining 18, eight used only graduate students or instructors in psychology and the other ten used college or normal-school students who were for the most part specializing in psychology. While this tendency to make use of adults, particularly of those adults immediately available for work as subjects in the psychological laboratory, is natural enough and has certain advantages (greater steadiness of application, ready grasp of instructions, possibility of illuminating introspection), it has likewise very definite disadvantages, as will be pointed out more fully farther on.

Similarly, there has been a decided tendency to carry on the investigation of the transfer of training with a quite limited number of subjects. Thus, in 13 of 29 investigations or sub-sections of investigations, conclusions of more or less import have been based upon the behavior of two, three, four, five or six subjects only. In 12 other investigations the number of subjects has ranged from 12 to 44.

Again, if the reports of 31 investigators whose work can thus be classified are grouped into laboratory investigations and schoolroom investigations, we discover that more than two-thirds of them (21 to be exact) are laboratory investigations.

We may say, then, that the typical experimental study of the transfer of training is the laboratory investigation in which the subjects are a small group of students who are specializing in psychology. It has long been apparent to many educational psychologists that conclusions reached by the laboratory study of the trained adult may not be too hastily generalized to apply to the mental processes of the growing child, and it is certainly wise to keep this caution in mind when dealing with such an issue as the possibilities and limitations of training, which, in the nature of the case, implies the engineering of modifications in the mind of a still immature organism.

III. Some Typical Investigations of Transfer and Their Results

1. Studies Dealing with Memory

(a) *James.* In his *Principles of Psychology* (1890), Vol. I, pp. 666-7, James reported the effect of training in memorizing a certain kind of material upon efficiency in memorizing other kinds of material. He was led to conclude that there was no transfer of training, that one's native retentiveness is unchangeable, that no amount of culture modifies a man's *general* retentiveness. This work of James represents, of course, a pioneer exploration. The experiment was loosely organized; in particular, as has already been noted, there was no control group.

(b) *Peterson.* The difficulty just mentioned in James' work was recognized in the repetition of the experiment in 1912 by Peterson,[3] but his data were too meager to permit reliable conclusions; in fact, three of the unpracticed subjects gained as much as one of the two practised subjects; nevertheless, we find Peterson saying that "there was certainly a large transfer of training here" (referring to one subject).

(c) *Ebert and Meumann.* The experiments of these investigators at the University of Zurich[4] on from 2 to 6 laboratory sub-

[3] Peterson, H. A. "Note on a retrial of Professor James' experiment on memory." *Psych. Rev.*, 19: Nov., 1912, 491-492.

[4] Ebert, E., und Meumann, E. "Ueber einige Grundfragen der Psychologie der Uebungsphänomene im Bereich des Gedächtnisses."

jects dealt with the effect of memorizing series of nonsense syllables upon efficiency in memorizing various other sorts of material. These experiments are of interest because the investigators claimed to have demonstrated the existence of a very large amount of transfer, explained not only as due to increased capacity to attend, better technique in learning, etc., but also as due to a hypothetical "sympathetic interaction of allied memory functions."

(d) *Dearborn.* The repetition of the work of Ebert and Meumann by W. F. Dearborn,[5] then of the University of Wisconsin, showed that a large portion of the transfer effect alleged by Ebert and Meumann was due to the gains made within the test series and that the remainder could be attributed to general improvement in orientation, attention, and better technique of learning.

(e) *Fracker.* A careful piece of work by Fracker,[6] from the laboratory of the University of Iowa, in 1908, dealt with the effect of training in remembering the order of sound intensities upon efficiency in remembering the order of various other materials. Fracker's results are regarded by him as furnishing clear evidence of transfer under certain conditions, provided "transfer" is thought of not as a mere "spread" of training, but as ability to use in a second situation a content or a form of procedure that is identical with the one trained. There is little spread of training without this identity of form or content. Ability to organize facts by conscious use of imagery, for example, is transferable, and it seems probable that it is especially in identical elements of an ideational sort that the carriers of the training effect are to be found.

(f) *Winch.* In 1908, and again in 1910, W. H. Winch,[7] of London, reported one of the first attempts to study transfer in children working under normal schoolroom conditions and with special pains to classify the subjects into two groups of equivalent ability. Winch's experimental groups were trained in rote memorizing of poetry or of meaningless things and the effect was meas-

[5] Dearborn, W. F. "The general effects of special practise in memory." *Psych. Bull.*, 6: 1909, 44.

[6] Fracker, G. C. "On the transference of training in memory." *Psych. Rev. Mon. Supp.*, No. 38, 1908, pp. 56-102.

[7] Winch, W. H. "The transfer of improvement in memory of school-children." *British Jour. Psych.*, 2: 1908, 284-293, and 3: 1910, 386-405.

ured that this training produced upon the memorizing of other forms of material. He construes the results as indicative of a real transfer of training; however, the amount of transfer cannot be regarded as great when proper allowance is made for the operation of chance and for certain features of method and material used in the study.

(g) *Sleight.* One year after Winch's second paper appeared another English investigation, that of Sleight,[8] in which 12-year-old school children and also adults in the laboratory were given training in memorizing poetry, arithmetical tables, and prose substance and were tested for the effect of this training on other forms of memorizing. The results give only slight ground for claiming transfer. The 'area' affected by special training is quite narrow, and if training does occur through the presence of identical elements, the identity must be very close and must include identity of procedure, of imagery, and of logical organization. In particular, the mere presence of an identical element is not sufficient to produce transfer; the element must also be usable by its possessor. Furthermore, a considerable degree of similarity between two bits of content may be entirely ineffective as a basis for transfer if there also exists an important difference in the two procedures. Again, there may appear actual interference with a latent tendency toward transfer if the second activity chances to be unpleasant for the subject. There is no "general memory faculty" and no generalized training of attention.

(h) *Woodrow.* Quite recently there has appeared a study of transfer of memory training, conducted by Woodrow[9] upon three groups of university sophomores. The 106 students in the Control Group took only the beginning and end tests; the 34 students in the Practice Group were given straight practice in memorizing poetry between the two tests; but the 42 students in the Training Group had, in addition, specific instruction in the technique of memorizing and in applying this technique. The second group (practiced) were sometimes better, sometimes poorer than the first

[8] Sleight, W. G. "Memory and formal training." *British Jour. Psych.*, 4: 1911, 386-457.

[9] Woodrow, H. "The effect of type of training upon transference." *Jour Educ. Psych.*, 18: March, 1927, 159-172.

group (control), but the third group (trained) average in the end tests 31.6 percent more gain than the control group. Woodrow says that "to determine the amount of transference of ability that did result from the study of a particular subject is one thing, to determine the amount of transfer which *might* be secured from the study of the same subject is a different and far more difficult thing." The experiment, he thinks, shows the difference "between unenlightened drill and intelligent teaching."

This study does seem to me especially interesting. It would seem to support the contention of those who ascribe particular significance to generalization as the vehicle of transfer. What appears to have happened is that the merely practiced students failed of themselves to educe adequate general principles governing the technique of memorizing, whereas these were supplied by the instructor to the Trained Group. On the other hand, the question may be raised as to whether we are really dealing with a problem of transfer (in the usual sense) if we simply demonstrate that direct instruction in a technique of intellectual procedure can afterwards be employed by intelligent mature students in various concrete situations. In other words, Woodrow should have shown us what amount of improvement could have been secured by instruction but no practice.

2. Studies Dealing with Perception, Discrimination, and Apprehension

(a) *Thorndike and Woodworth.* In their laboratory experiments reported in 1901 these authors[10] trained from four to six laboratory subjects in estimating small areas, weights or lines and in the perception of certain letters and parts of speech, then measured the effect of this training upon the estimation of other areas, weights, and lines, and the perceiving of other letters and parts of speech. Despite the close similarity of the training and the test series, the general quantitative outcome indicated a comparatively slight transfer of training, and there were evidences of interference (negative transfer) under certain conditions. The authors

[10] Thorndike, E. L., and Woodworth, R. S. "The influence of improvement in one mental function upon the efficiency of other functions." *Psych. Rev.*, 8: 1901, 247-261, 384-395, 553-564.

of this often-quoted study contend that there are no such "general" functions open to training as attention, accuracy, or power of observation, and that whatever transfer takes places does so by dint of the presence of identical elements in the influencing and influenced functions.

My criticism of this contention has already been foreshadowed. Why expect academically trained adults to be open to further improvement of such "general functions"?

(b) *Judd.* While working in the Yale laboratory, Judd[11] trained two subjects in the Müller-Lyer illusion and found that, by dint of practice, and without the aid of abstract judgment, the illusion gradually disappeared and that the practice-effect was transferred. The effect was ascribed to perceptual training.

(c) *Coover and Angell.* Six subjects trained in introspection were used by Coover and Angell[12] in their investigation of transfer in the field of sensory discrimination. The effect of training in the discrimination of sound intensities was found to transfer to the discrimination of grays, but this result is supported by assertions based upon introspection rather than by any clear statistical demonstration, since the two control subjects seemed to improve as much as the four experimental subjects.

(d) *Whipple,*[13] *Foster,*[14] *and Dallenbach.*[15] The experimental work of these three, all studies from the Cornell Educational Laboratory, may be presented conjointly. In all three investigations the material consisted of drawings, numbers, lines of poetry, groups of objects, and the like, exposed to view very briefly.

The work of Whipple and of Foster was limited to laboratory test upon competent adults. The general outcome was that there was no improvement of any basic intrinsic ability, like attention or apprehension or visualizing, that the superficial improvement

[11] Judd, C. H. "Practice and its effect on the perception of illusions." *Psych. Rev.*, 9: 1902, 27-39.

[12] Coover, J. E., and Angell, F. "General practice effect of special exercise." *Amer. Jour. Psych.*, 18: 1907, 327-340.

[13] Whipple, G. M. "The effect of practice upon the range of visual attention and of visual apprehension." *Jour. Educ. Psych.*, 1: 1910, 249-262.

[14] Foster, W. S. "The effect of practice upon visualizing and upon the reproduction of visual impressions." *Jour. Educ. Psych.*, 2:1911, 11-21.

[15] Dallenbach, K. M. "The effect of practice upon visual apprehension." *Jour. Educ. Psych.*, 5: 1914, 321-334, 387-404.

was very specific and explicable as due to habituation to the setting of the experiment, to special 'tricks' or techniques rapidly acquired during the experimenting, to the development of confidence, etc.

In the work of Dallenbach, however, when similar rapid-exposure practice was given to a class of third-grade school children, not only was there a marked practice-effect, but also evidence of the transfer of this effect, brought out more especially by retesting after 11 months in comparison with untested pupils. Dallenbach concluded that the training had permanently altered and modified certain mental traits; in other words, that such exercises as Whipple and Foster found to be ineffective in altering the basic capacities of adults were really effective in making such alteration in children.

(e) *Ruger.* Either in this section or the next may be classed the investigation conducted by Ruger[16] at Columbia University. By employing 37 mechanical puzzles and by minute records of the subjects' work, Ruger was able to observe to what extent solutions arrived at in given puzzles were transferred to the solution of other puzzles of varying degrees of similarity. The outcome of this study seems quite unlike that reported by Judd in the practice with illusions of length, for Ruger concluded that the presence of imagery was of no avail without conscious generalization of methods of attack and analysis. Other factors making for transfer were ideals of efficiency, attitudes of "high-level attention," satisfaction of success, etc. "The value of specific habits under a change of conditions depended directly on the presence of a general idea which would serve for their control."

(f) *Wang.* Somewhat similar to the study of Coover and Angell is that carried on at the University of Michigan by a Chinese student, Wang,[17] who trained a small number of pupils in discriminating the lengths of vertical lines and tested the effect of this training upon their ability to discriminate pitch, to discriminate shades of color, to discriminate sizes of figures, and to mark words containing certain letters. Wang found that no trans-

[16] Ruger, H. A. *The Psychology of Efficiency.* Archives of Psychology, No. 15. 1910.

[17] Wang, C. P. *The General Value of Visual Sense Training in Children.* Educ. Psych. Monographs, No. 15, 1916.

fer appeared unless the children were able to develop an efficient method in the training series and to use this purposefully in the test series.

3. Studies Dealing With Voluntary Effort and Motor Adjustments

(a) *Gilbert and Fracker.* One of the earliest studies in this field is that of these two Iowa investigators,[18] who worked in the laboratory with the stock forms of reaction-time. For their three subjects large amounts of transfer-effect were noted; thus, practice in simple sound reaction reduced the time for simple reaction to other stimuli almost as much as it reduced the time for the sound reaction; similarly with practice in discrimination reactions, but practice with the former did not affect ability in the latter. It must be evident that the general setting of the reaction test is so similar for the various types of reaction that one would expect such an improvement as was reported to occur through such external or indirect factors as habituation, acquiring favorable set, etc.

(b) *Bair.* In the Columbia Laboratory, in 1902, J. H. Bair,[19] in connection with various experiments on the practice curve, encountered situations permitting a study of transfer, *e.g.*, the effect that training in tapping rapidly on the typewriter a given series of exposed characters had upon speed of tapping other series, and again, the effect that training in reciting the alphabet rapidly under certain conditions had upon reciting it under other conditions. The general outcome of Bair's work was evidence for a definitely transferred practice-effect; "one kind of practice helps us in another kind." Bair also reported numerous interference effects, and it is of interest to note that interference often lessens as practice continues, until it disappears and the transfer may then become positive and favorable to the new activity.

(c) *Coover and Angell.* In the investigation by these authors that has already been mentioned an experiment was included in

---

[18] Gilbert, J. A., and Fracker, G. C. "The effects of practice in reaction and discrimination for sound upon the time of reaction and discrimination of other forms of stimuli." *Univ. of Iowa Studies in Psych.*, 1: 1897, 62-76.

[19] Bair, J. H. *The Practice Curve.* Psych. Rev. Mon. Supp., No. 19. 1902.

which subjects were trained for forty days in the rapid sorting of cards. This training was said to have definitely increased their efficiency in certain aspects of typing. The investigators designed this experiment to eliminate all identical factors and hence ascribe the statistical improvement to certain "general" factors, like equitable distribution of attention and development of power to concentrate attention throughout an entire series of reactions. This is one of the few studies in which a claim has been advanced for such training of a general function.

(d) *Wallin.* The experiment of Wallin,[20] in which two subjects were practiced for 60 days in controlling with one eye illusions of reversible perspective with the result that this practice was found to be transferred to the other eye, may be cited as a typical instance of the so-called "cross-transfer of training." It seems obvious that the training occurred primarily elsewhere than in the used retina, so that its effect would naturally appear when the previously unused retina was employed. The same remarks may be made concerning Wallin's further observation of transfer from foveal to peripheral perception of these illusions.

4. Studies Dealing With Schoolroom Activities and Attitudes

(a) *Bagley, Squire, and Ruediger.* An experiment planned by Mrs. Carrie Squire and reported by Bagley[21] concerned the transfer of a habit or ideal of neatness from one branch of school work to other branches. This led Ruediger[22] to investigate the matter further a few years later.

The Bagley-Squire work seems to have been seized upon as a basis for pedagogical generalizations that far exceed its warrant. In substance "the test consisted simply in insisting on neatness and accuracy in the preparation of arithmetic papers by pupils in the latter half of the third grade." Three weeks' persistent drill produced definite results in these papers, but no results, save a slight

[20] Wallin, J. E. W. "Two neglected instances of the transfer of training." *Jour. Educ. Psych.*, 1: 1910, 168-171.

[21] Briefly in his *Educative Process* (1905), p. 208, and more fully in his *Educational Values* (1911), pp. 188-189.

[22] Ruediger, W. C. "The indirect improvement of mental functions through ideals." *Educ. Rev.*, 36: 1908, 364-371.

loss, in improving neatness in the language or spelling papers. The contrast was "almost startling."[23] Drill in specific habits of neatness did not produce a general habit of neatness.

In Ruediger's subsequent extension of the neatness experiment the specific training in neatness in one school subject was supplemented by talks on neatness elsewhere (in dress, in the home, etc.). Small gains appeared in neatness, not only within the school subject specifically trained but also within other subjects. Ruediger stresses in especial the operation of conscious ideals as an important factor in effecting the transfer of the habit.

(b) *Winch.* In addition to the investigation previously mentioned, this British investigator[24] undertook several studies of the transfer of training in numerical accuracy in school children. In the first set of experiments, of the four studied, there was no transfer in one school, but definite transfer in the other three schools. The transfer here consisted of the effect of training "in working rule sums" upon subsequent accuracy in solving arithmetical problems. In the later study, in which 36 ten-year-old boys were given ten practice exercises in numerical computation, this group did not surpass the unpracticed control group of 36 boys in subsequent tests of ability to solve arithmetical problems rationally (but without actual numerical calculations), that is, drill in computation did not transfer to skill in problem solution. Of interest in this investigation is Winch's demonstration that the existence of a high positive correlation (0.85 in one case) between these two abilities does not permit us to argue to a causal relationship such that the improvement of one of the related abilities necessarily produces improvement of the other.

---

[23] This phrase was used by Bagley in 1905. He then said: "*Careful* experiments were undertaken to determine whether the habit of producing neat papers in arithmetic will function with reference to neat written work in other studies." But in 1911 he wrote: "Experiments undertaken in the Montana State Normal College *seemed* to substantiate these conclusions. These experiments tested *very crudely* the ability to transfer the results of training in neatness and accuracy" (italics mine). It is this later statement which suggests that rather too much significance may have been attached to the Bagley-Squire experiment.

[24] Winch, W. H. "Accuracy in school children. Does improvement in numerical accuracy 'transfer?'" *Jour. Educ. Psych.*, 1: 1910, 557-589. Also "Further work on numerical accuracy in school children. Does improvement in numerical accuracy transfer?" *Ibid.*, 2: 1911, 262-271.

(c) *Starch.* Though carried on (presumably) in the laboratory with college students, we may class among the studies of school activities the study reported from the University of Wisconsin by Starch,[25] in which eight subjects were drilled for 14 days on mental multiplications and in which their efficiency in subsequent tests of various arithmetical operations and in memory span for words was compared with that of an unpracticed control group. There was "little change for memory span for either group," but "the practiced observers showed from 20 to 40 percent more improvement in the arithmetical tests than the unpracticed observers." In weighing the significance of this large transfer-effect, however, we should note that the abilities demanded in the test series were closely similar, if not at times almost identical, with those demanded in the drill series.

(d) *Wallin.* Another school subject, that of spelling, was the material chosen by Wallin[26] for a study of transfer of efficiency gained in 'column drills' to performance in dictated compositions. This investigation was made on more than a thousand pupils of the public schools of Cleveland. Some three to thirteen weeks after several weeks of drill in column spelling, the pupils were graded on their spelling of 40 to 50 words employed by their teachers in dictation. Wallin concludes that the column drill may produce a positive increment of spelling efficiency, an increased facility to master new forms. Despite the large number of pupils employed in this study, the experimental conditions were not ideal to yield clear-cut evidence of transfer (there was, for instance, a certain amount of incidental teaching of spelling).

(e) *Briggs.* Decidedly complicated is the contribution of Briggs,[27] who employed two groups of 7th-grade children for an investigation of the effects of formal instruction in English grammar. An extended series of 54 problems and exercises constituted the preliminary and final test series and the intervening drill was in grammar taught somewhat inductively. Supplementary trials

[25] Starch, D. "Transfer of training in arithmetical operations." *Jour. Educ. Psych.*, 2: 1911, 306-310.

[26] Wallin, J. E. W. *Spelling Efficiency in Its Relation to Age, Sex, and Grade.* Educ. Psych. Monographs, 1911.

[27] Briggs, T. H. Formal English grammar as a discipline." *Teachers Coll. Record,* 14: Sept., 1913.

of the tests were made in five public school systems, some of which taught grammar and some of which did not. Taking the average performance of the classes under test, the net outcome is negative, *i.e.*, to quote Briggs: "It may safely be asserted that these particular children, after the amount of formal grammar they had, do not, as measured by the means employed, show in any of the abilities tested, improvement that may be attributed to their training in formal grammar."[28]

(f) *Hewins.* The monograph by Miss Hewins[29] deals with a more mature group of pupils (high-school freshmen) and with a different mental function (training in observation). After dividing the pupils of several classes into equivalent groups on the basis of suitable preliminary tests, one group in each class was drilled on the observation of biological material. Final tests were given to the entire class on the observation of non-biological material. On the whole, the results justify the assertion that "there are certainly conclusive evidences of transfer."

(g) *Rugg.* Omitting detailed reference to Bennett's attempt[30] by studying college marks and scores in laboratory tests to show that the four years of college work is effective only in increasing specific abilities along the lines of special interests, we may cite the work of Rugg[31] as perhaps the most comprehensive study of the transfer of training among college students. Five hundred University of Illinois students were tested, of whom 326, the training group, pursued a course in descriptive geometry. The aim of the investigation was to determine the extent to which the special training received in this course operated to increase efficiency in various other mental activities of different degrees of similarity to the activities trained. The results show that the practice in geometric visualization received in the 15 weeks' work in descriptive geometry did substantially increase the efficiency of "mental manipulation" of both geometric and non-geometric elements, and

[28] The reader interested in a critique of this investigation may consult a review of it by Bagley, *Jour. Educ. Psych.*, 5: 1914, 538-540.

[29] Hewins, Nellie P. *The Doctrine of Formal Discipline in the Light of Experimental Investigation.* Educ. Psych. Monographs, No. 16, 1916.

[30] Bennett, C. J. C. *Formal Discipline.* Columbia University Contributions to Education, 1907.

[31] Rugg, H. O. *The Experimental Determination of Mental Discipline in School Studies.* Educ. Psych. Monographs, No. 17, 1916.

that the degree of transfer, as would be expected, varied with the degree of similarity between the practise and the test activity. As for the agencies of transfer, these were found by Rugg to lie not only in the direct carrying over of specific habits, but also in better methods of attack, better attitudes of orientation, and an increased facility in manipulating many elements at the same time.

(h) *Thorndike.* One of the most ambitious recent studies of transfer effect in the school is that reported by Thorndike[32] who used statistical methods applied to intelligence test scores to determine the effect of a year's school work in various high-school subjects upon the amount of the traits measured by these test scores. The essentially negligible disciplinary effect thus unearthed may perhaps be set forth most effectively by the following quotation:

"If our inquiry had been carried out by a psychologist from Mars, who knew nothing of theories of mental discipline, and simply tried to answer the question 'What are the amounts of influence of sex, race, age, amount of ability, and studies taken, upon the gain made during the year in power to think, or intellect, or whatever our stock intelligence tests measure,' he might even dismiss 'studies taken' with the comment: 'The differences are so small and the unreliabilities are relatively so large that this factor seems unimportant.' The one causal factor which he would be sure was at work would be the intellect already existent. Those who have the most to begin with gain the most during the year. Whatever studies they take will seem to produce large gains in intellect."

Special mention may be made also of the conclusion that the physical sciences are equal, if not superior, to languages and mathematics in respect of mental discipline from any point of view.

IV. What Has Been Proved Concerning the Amount of Transfer

While by no means exhaustive, the foregoing description of typical investigations has been sufficient to indicate the characteristic methods by which the problem of transfer has been attacked experimentally. But the outcomes of these investigations are so varied that some attempt should be made to summarize their general trend. This I shall attempt by setting down a series of categorical statements.

---

[32] Thorndike, E. L. "Mental discipline in high-school studies." *Jour. Educ. Psych.*, 15: 1924, 1-22, 83-98.

1. *In the light of what we now know to be necessary for reliable inference, many of the experimental studies of transfer are so inadequate in scope and method that they afford altogether too shaky ground for establishing permanent conclusions.*

Thus, the subjects have been too few in number or already too well 'crystallized' mentally to expect any marked transfer of improvement. The material upon which the training has been accomplished has often been, I think, too 'picayunish' to exert any natural appeal. Again, there has been often no control group or there has been no effort to compute statistically the extent to which mere chance has affected the result. Yet again, the training has been too brief to promise any far-reaching or significant alteration of the mental efficiency of the recipient. Finally, many of the investigations have picked for their subjects a half-dozen graduate students or instructors, who, in the very nature of the case, represent a type of mind that has been rigorously selected for having already attained a superior stage of mental training. No doubt these objections may be exaggerated, but personally, I cannot avoid the conclusion that a considerable part of the experimental work on the transfer of training might as well be thrown out of the window bodily so far as any significant contribution may be expected to our knowledge of the real possibilities and limitations of training.

2. *There can be counted on the fingers of one hand the experimental studies that have definitely yielded the verdict: "No transfer."*

Here might be placed, for instance, the pioneer experiment of James, the neatness experiment of Mrs. Squire, the correlation study of Bennett, and the formal grammar study of Briggs. Some of the studies that have been quoted as supporting this position, like my own study of visual apprehension, are not properly classed as studies of transfer. They have merely shown certain essential limitations in the nature of the practice improvement that occurs in adults when faced with the acquisition of a new set of responses. Thus, to deny that practice in rapid observation improves any intrinsic capacity or faculty of observation itself is not to deny the possibility of the transfer of training. It is unfortunate that this interpretation has gained such currency among certain popularizers of the experimental literature.

3. *Several studies in which the investigator has claimed a large or a very definite amount of transfer provide inadequate proof of this claim or deal with a type of transfer in which the shift from trained to untrained activity is so slight as to be of little practical significance.*

In making the assertion of "inadequate proof," I have in mind the investigations of Ebert and Meumann and of Coover and Angell. In making the assertion of "no practical significance" I have in mind the so-called "transfer of training" in the reaction-time tests of Gilbert and Fracker or Wallin's 'other eye' transfer or Wallin's transfer from spelling in a column to spelling in a sentence. To exaggerate to the point of grotesqueness, of what avail to demonstrate that having trained myself to sing a song while wearing a hat, I find I have improved in singing it with my hat off?

4. *In a preponderant number of investigations there has been revealed a measurable, though restricted amount of transfer. In general, the more similar the influencing and the influenced activities, the more certain the transfer of efficiency and the greater its amount.*

Anyone who reads the literature of formal discipline in chronological order will perceive clearly that investigators have been swayed at times by their own attitude toward the popular notions of transfer prevalent when their investigations were conceived. The effect of this attitude is not infrequently revealed in the manner in which the investigator has formulated his conclusions. If he thinks there is current a naive belief in general cultural training, in wide-spread transfer, then a definite, though moderate amount of transfer in his experiments is set forth as a demonstration of the specificity of training and the futility of belief in formal discipline, although it might quite as well have been heralded as a partial justification of the dogma.

5. *The transfer of training may impair the second activity, rather than improve it. It appears likely that this so-called 'negative' transfer is most pronounced when the second activity is begun and that it may not only be reduced to zero, but become transformed into a positive transfer if the second activity is continued long enough.*

This outcome is perhaps sufficiently perplexing at first view to warrant a simple illustration. A man well trained in operating a standard automobile gear-shift, when suddenly called upon to operate a car with reversed shift (like the Dodge, for instance), might, on account of interference, perform the shifting at first less efficiently than a novice; if he persisted, however, he would eventually surpass the novice, because he would already have learned the underlying general principles of correct gear-shifting and these would be 'carried over' advantageously, once the specific differences between the two cars were overcome.

V. What Has Been Proved Concerning the Agencies of Transfer

1. The Doctrine of 'Common Elements'

Referring to our schematic explanation of the problem of transfer and to the summary just made of the quantitative results of transfer experiments, it is evident that there can be neither complete specificity of response nor complete interlocking of all responses, because there is commonly discovered a certain amount, but a restricted amount, of influence from training series to subsequent test series. And wherever there is spread, or transfer, of improvement from one activity to another, it is certainly a fair presumption that the two activities—neural activities, of course, in the last resort—must have one or more features in common.[33] This argument is equivalent to declaring that transfer takes place through "identical elements." If this phrase—"common elements"—be taken in a broad sense, all the theories of transfer, so far as I can see, can be said to agree, that is, all explanations of transfer posit the operation of a common factor between the influencing and the influenced function. The point of dispute lies, however, in the question: What are the common elements that are

[33] Poffenberger, in summarizing his study on "The influence of improvement in one simple mental process upon other related processes" (*Jour. Educ. Psych.*, 6: 1915, 459-474), concludes similarly: (1) that if there are no identical associative processes common to two activities, training in the first activity has no influence at all on the second, (2) that if there are some identical processes, there will be positive transfer, but (3) that if the second activity necessitates breaking associative connections learned in the first, *e.g.*, adding after multiplying, there will be negative transfer, or interference.

ordinarily responsible for transfer? To examine this issue, a distinction, useful for purposes of discussion, may be drawn between common content and common method.

2. Common Content

So far as common content is concerned, this type of transfer may be readily illustrated if we think of 'content' as 'items of information.' Thus, the study of Latin helps the student later with his French because the French is a Romance tongue and many French words have been carried over with slight modifications from the Latin. It sometimes seems as if this sort of carrying over of learning were too obvious or commonplace to be dignified by the designation "transfer." I mean there hardly seems to be any real transfer. You perceive readily that the French *terre* must mean *earth* because you already know that the Latin *terra* means *earth*. Yet this sort of carry-over is commonly cited as a case of transfer of training because the *terra-earth* response was acquired in one situation (translating Latin) and now appears as an aid to the slightly variant *terre-earth* response in another situation (translating French). Or, to use another illustration, we are discussing identity of content when we raise such a question as: "Does training in educational psychology given to prospective teachers in the school of education transfer to their classroom teaching?" because we mean: "When a concrete situation arises wherein a certain psychological precept properly applies, does the teacher recall the precept, see and undertake its application?" The same aspect of the general problem appears, to cite another instance, when we inquire whether training in the responses of the multiplication table transfers adequately to the use of multiplication in problem solving.

3. Common Method

If by 'method' we understand 'way of going about doing something,' in the broadest sense of that phrase, it seems to me that all transfer that does not occur from community of content must occur from community of method. But we find that the literature teems with lengthy discussions of the possibility or impossibility of transfer through *generalized method* of various sorts, which is by

many thought of as distinct from transfer through what they term "common procedure." I judge that a few concrete illustrations may be helpful in setting forth this controversy by seeking what distinction, if any, exists between common procedure and generalized method.

(a) *Common Procedure.* Suppose that a child discovers, when being trained to commit to memory the outline of a map, that it pays him to close his eyes and attempt to see the map in his mind's eye. Suppose that, subsequently, when being tested in committing to memory a list of words to be spelled, he employs to advantage the same method of closing his eyes and trying to visualize each word. As I understand it, most investigators would agree that in this case transfer of training takes place through identity of procedure.

(b) *Generalized Method.* Suppose that the same child, while being trained to observe a card of letters that is flashed before his eyes for a fraction of a second, discovers that it pays him to concentrate his attention upon a limited number of the letters and make sure of these, rather than to try to take in every letter that is exposed. Suppose that, subsequently, when being tested in recalling the gist of a rather long story that is told him, he recalls his previous experience with the card of letters and definitely resolves to concentrate his attention upon a few fundamental features in the story and not try to remember the remainder. I cannot see that the transfer in this second case differs in principle from that involved in the map and spelling case, yet we find that the discovery and use of a "method of distributing the attention" is classed by some writers as "transfer by generalization," and as something that differs from "identity in procedure."

Suppose, thirdly, that this child, though somewhat dismayed by the card exposure training at first, was greatly cheered by learning that he was really doing better than his mates, so that he said to himself: "I can do this better than the other boys if I only try hard." And suppose that this same encouraging thought stimulated his work advantageously later when he had to recall the story. It seems to me that we may say that, in the broad sense of the term, we have here again a case of transfer through 'method,' through the development of a 'way of going about it,' but most

investigators feel that this type of transfer effect must have a special category; it is called "transfer through the agency of a generalized attitudinal factor." Bagley, as I understand him, might class this illustration as "transfer through ideals."

Perhaps these three illustrations will suffice to make clear my point, namely: *Unless there is actual partial identity of subject matter, or content, any transfer of training that appears is the result of the presence of a common factor of method.*

VI. The Pedagogical Significance of Transfer

1. Conscious *Versus* Unconscious Transfer

The experimental work has made it evident that efficiency acquired in training may spread to related activities without any deliberate intent on the part of the person trained and in fact without his knowing what increased his efficiency or that any such factor was operating in the related activities. Perhaps it would be better to term this "unintentional" or "passive" or "automatic," rather than "unconscious" transfer. If it were not for this sort of transfer, the limitations of training would be far more serious than they are.

But the experimental work has also made it evident that in many situations (and I think especially in transfer through community of method) transfer is either non-existent or decidedly limited unless the factors that are producing improvement in the training operations are consciously recognized and also consciously applied to bring about efficiency in the related operations. This is what Sleight refers to, for instance, when he says that mere presence of an identical element is not enough; it must also be usable by its possessor. The pedagogical moral is obvious: *Teachers should arrange the work of pupils and their own instructional efforts in such a way as to facilitate the conscious recognition by the pupils of the methods by which efficient mental work is done.*

2. Facts *Versus* Relations as Vehicles of Transfer

Another way of bringing out the relative importance of community of content and community of method in effecting the transfer of training is to compare the value of facts with the value of

relations as vehicles in the process. When two or more facts are being handled in thinking, it is the relation of these facts to one another and the similarity of this relation to other relations between facts previously handled that furnishes the usual basis for whatever transfer of training occurs. It is a relation that the individual has educed from consideration of the first set of facts that enables him to handle more efficiently the second set of facts.[34]

The pedagogical moral here is likewise obvious: *A skillful and 'artistic' teacher will take pains to see that facts are so presented that pupils draw out the relations between these facts and utilize these relations in the handling of other sets of facts in which similar relations obtain.*

This leads us naturally to another important pedagogical principle.

3. Dependence of Transfer Upon Native Mental Ability

If efficient transfer hinges primarily upon the conscious realization and conscious utilization of relations between facts and relations between relations, it can hardly be denied that the possibilities of transfer are to an important degree dependent upon what we commonly term "general intelligence," upon native mental ability in intellectual operations. It is not necessary at this point to drag the reader into the controversy concerning the fundamental nature of intelligence, for I am sure all those who have discussed it are willing to agree that superior intelligence is not primarily conditioned by superior operation of the simple processes of sensation, perception, or even of the processes of sensory discrimination or imaging or memory, but rather by the superior operation of the relational processes. Put in other words, the outstanding feature of dullness is the comparative limitation of mental processes to sensitivity, bare retentivity, and simple associative reproduction,

---

[34] Some writers go so far as to say that no transfer takes place at all from the bare perception or cognition of facts, but only with respect to the apprehension and use of relations educed between sets of facts (see, for instance, J. J. Strasheim, *A New Method of Mental Testing*, p. 147, and authorities quoted by him, like G. Thomson, *Instinct, Intelligence, and Character*, pp. 140 ff.). The eduction of such relations may be extended, of course, to higher and higher levels; there may be relations between relations, and these "higher level relations" may extend, in competent minds, to almost any level of generalization.

whereas the outstanding feature of brightness is the comparative predominance of the apprehension, formulation, and utilization of generalized relations in the intellectual life.[35] Methods, laws, rules, maxims, precepts, are embodiments of just such generalized relations. Whenever the experimental conditions are so arranged as to permit the possession of these outstanding features to emerge (for example, in the investigation of Strasheim, just mentioned), the contrast between dull children and bright children is striking, and the dependence of transfer upon native ability is clearly demonstrated.

The pedagogical moral is again obvious: *Teachers may expect bright children to surpass dull children in the amount of transfer gained from specific training; moreover, in proportion as the child's intelligence is inadequate, in that proportion must the teacher himself consciously and purposefully bring the child's attention to the relations to be educed and to the use of these relations subsequently. Even when aided by such skillful instruction, the spread of training will be decidedly less in the dull.*

4. Dependence of Transfer Upon Maturity

Experimental investigation makes it evident that, as the child passes from infancy to maturity physically, there occurs a roughly similar growth in numerous mental abilities. This growth may be regarded, at least from the standpoint of theory, as a simple process of maturing, as a development that would take place with the lapse of time whether there were any formal training or not.[36] Acceptance of this doctrine is implicit, for instance, in all our machinery for computing mental age and intelligence quotients. From this point of view, we expect that the ability to educe and apply relations will increase with maturity from birth to some period in the neighborhood of 14 years, and consequently, we would expect a growth during childhood and youth in the extent to which specific training can become generalized.

---

[35] On the testimony of teachers to this effect, see Baker, H. J. *Characteristic Differences in Bright and Dull Children.* Bloomington, Illinois, 1927.

[36] Of course, other definitions of maturity may be defended, but this one is at least useful for schematizing the factors underlying mental development, as I have tried to show elsewhere ("Endowment, maturity, and training as factors in intelligence scores," *Scientific Mo.*, 18: 1924, 496-507.)

But, if we leave the matter here, we should expect that from the age of adolescence to perhaps some point where the processes of senescence intervene, the capacity to receive training and to profit by its spread would remain constant and maximum. Just here, however, we are confronted by considerable experimental evidence that the spread of training in adults is not large. The solution of this seeming contradiction appears to me quite simple, namely: by the time the average individual attains mental maturity a considerable fraction of the mental operations that were originally susceptible to modification by training have already been modified and have become more or less definitely 'set' in a fashion that has turned out to be satisfactory as far as the individual's daily needs are concerned. Thus, why expect training in sensory discrimination of college students to increase their fundamental capacity to attend to small differences? They have been attending to such differences, so far as life necessitated it, daily for twenty years; they are already trained, by hours and hours of mental work, to concentrate attention, to disregard distractions, to formulate their experiences in fitting terms, and in general, to exert a reasonably efficient control over their intellectual processes. No, the place to look for striking transfer-effects is in the immature. To my mind, this argument is well illustrated by the marked difference in outcome between the experiments of Dallenbach and those of Foster and myself. Neither Foster nor I could discern any marked intrinsic improvement in adults (mostly highly selected adults) from our training in 'observation' under conditions of rapid exposure, but Dallenbach found good evidence of a fundamental and permanent improvement in this function when he applied an almost identical process of training to third-grade school children.

From these considerations comes the conclusion: *The possibilities of the transfer of training should be found at their maximum in children, whose mental activities are still in the formative stage. While adults possess a much higher intrinsic capacity than do children to effect the eduction, the conceptualization, and the applications of relationships, this capacity is less evident in practice simply because the vast majority of these relations have already been educed, conceptualized, and applied under the stress of the extended daily experience and training they have already had.*

5. Obscurity of the Higher Relations

Transfer of training through a common content, as we have already pointed out, is so obvious as hardly to merit the name of transfer. But transfer through community of method, when 'method' has reference to the operation of relations between relations, of the 'higher level relations,' is often exceedingly obscure. Right here is, indeed, the basis for the never-ending arguments about the cultural value of the classics and other branches of subject matter that have long been favored for their merits as producers of 'formal discipline.' Take the case of the advocacy of mathematics as a trainer of reasoning. The experimentalist who brushes this claim aside with the assertion that "all training is specific" and that "there is no such faculty as reasoning" must at times feel a qualm of conscience and wonder if that line of retort may not, after all, "pour out the baby with the bath." Take the case of the pleader for Latin who declares that drill in Latin grammar teaches the student "the importance of little things" (meaning on account of case endings and the like). I hold no brief for Latin, nor would I ever advocate the study of its grammar as a valuable way to learn the importance of the seemingly trivial, but I am concerned to point out that what this pleader declares to happen really *could* happen. I mean that it is within the bounds of possibility that a Latin student should educe a generalization from this study that he never chanced to educe before—"I must pay more attention to little things; they sometimes make all the difference in the world"—and within the bounds of possibility that this student should remind himself of his generalization ten years later when he was called upon to judge the merits of rival candidates for a political office. In other words, reverting to the phraseology of my introductory paragraphs, no one can assert positively that the training received in acquiring Response 34 may not help years later in effecting Response 635,782, if 34 and 635,782 happen to have some community (even a very tenuous one) in X. But, of course, it is equally true that to locate and isolate X and to prove that it *was* common to both Response 34 and Response 635,782 and further to prove that it was on account of this community and *by means* of it that the latter response was rendered more efficient is, indeed, exceedingly difficult. Let the reader consider some one

ability of which he is a bit proud, and let him ask himself just where in his early training, in what classroom, by what teacher, in what lesson, or in what experience or series of experiences was laid the foundation of this ability; if he is candid, he will almost certainly have to answer: "I don't know." If we are to make headway toward a final formulation of transfer, future investigations will need to be directed toward the careful analysis of these obscure, higher-level intellectual operations, the genesis and function of various attitudes and ideals, the development and use of the larger generalizations of method—an exceedingly difficult field for study.

My argument here may be stated, therefore, more briefly: *It is not only probable, but fairly certain, that some of the most important agencies of transfer are to be found among the higher-level relations, in generalized attitudes, moods, ideals, sets, ways of going about mental operations generally. These agencies are in the nature of the case obscure; the nature and extent of their participation has occasioned extended argument, largely futile because so little is known factually. Further study is much needed, but difficult. In the meantime the educator is not justified in resorting to any specific subject of instruction for the purpose primarily of deriving from it indirect training values, but neither is he justified in neglecting to derive from every subject all the training value that it seems to promise.*

6. Moral Training

The experimental study of the transfer of training has dealt almost entirely, as we have shown, with comparatively simple mental operations of the intellectual type, isolated for laboratory control and inspection, as, for instance, with rote memory, pitch discrimination, speed of adding, estimation of size, etc. Nevertheless, in discussing the agencies of transfer we find mention made in these studies, as well as in studies of more complex schoolroom activities, of the operations of certain general factors, such as confidence, persistence, ideals of accuracy, ideals of neatness, that may be regarded as quasi-moral, if not moral traits. The claims made by proponents of the cultural merits of the classics likewise may be regarded as evidence of a belief—though not, of course, a demon-

stration—that a transfer of training exists within the realm of moral development. We find on all sides this conviction that moral training is gained by transfer. The playground inculcates the spirit of "fair play." How? Surely, if at all, by a transfer of training just as real as the transfer of training between any purely intellectual activities. The point I wish to make is that a very considerable reliance is placed upon the indirect approach in the development of character. I am almost tempted to say that an unquestioned acceptance of the doctrine of formal discipline, of the thoroughgoing efficacy of transfer of training, is a conspicuous feature of plans for moral training. Opinions will differ as to the strength of this foundation for moral development; to some psychologists at least, transfer seems an uncertain and insecure reliance for this highly important aspect of education. Whatever be the merits of the case, it is certain that we know less about the amount of transfer and the agencies of transfer in the realm of moral, than in that of intellectual training, despite the fact that from a practical point of view it is just in the realm of moral training that we should profit most by full knowledge of the possibilities and limitations of training. To conclude this final comment on the pedagogical significance of transfer:

It is particularly in the field of moral education that there has prevailed a general and uncritical acceptance of the dogma of formal discipline. While it is probably true that through the agency of the higher-level relations there may occur an important development of certain quasi-moral attitudinal traits, there has been no clear experimental demonstration of this transfer and it remains precisely in this important field of moral education that we know least about the possibilities or the limitations of training.

PREFATORY NOTE TO CHAPTER XIV

When mental tests first were used, the opinion was frequently expressed that the large individual differences brought to light were perhaps due chiefly to the unequal familiarity of the examinees with the tasks set, and that among examinees who had been practiced well toward their limit these differences would become small or would disappear. In that event mental-test scores of unpractised subjects would lose their significance entirely. If individual differences can be eliminated by training, then they are presumably largely due to environmental factors. If they persist, in spite of equalized training, they are presumably to be attributed to nature rather than nurture. That is the problem with which the following chapter is concerned.

The authors, however, have not investigated the effect of practice upon individual differences in mental-test scores, but instead, its effect upon individual differences in initial learning scores. It is probable that whatever holds true for the latter comparison holds also for the former. If different subjects do not maintain much the same relative position in initial and final stages of practice in mental or psycho-motor performances, a serious disturbing factor enters into all our intelligence and personality tests.

The issue is complicated by the probability that what holds for one kind of test or learning material may not hold for another kind. It may be that practice decreases individual differences in simple psychomotor performances, but increases them in performances which tax the higher mental processes. Extensive investigations will be necessary to determine the exact facts. The data presented in this chapter are interesting, but are not regarded by the authors as at all conclusive. The most important result is that, in digit-symbol substitution, card-sorting, mental multiplication, and achievement-test performances, the examineers tend to keep much their same relative position at different stages of practice. The individual differences here investigated are not greatly altered by equalized training.

CHAPTER XIV

THE EFFECTS OF PRACTICE ON INDIVIDUAL DIFFERENCES[1]

JOSEPH PETERSON
George Peabody College for Teachers, Nashville, Tennessee
AND
MYRON CALL BARLOW
University of Utah, Salt Lake City, Utah

NATURE OF THE PROBLEM

Does practice in any mental function or any specific kind of performance increase or decrease the individual differences present at the beginning of the practice? Again, do the individuals that are least efficient at the beginning tend to be least efficient throughout the practice or are there frequent changes in relative rank in efficiency during the practice or other similar training? These are questions of considerable importance both to educational work and to vocational direction, and they have received a fair amount of attention by experimental psychologists. But the first question is yet in the main to be answered.

As to the answer to the second problem as here stated there is now rather general agreement. Numerous investigators, concerned with various kinds of performances, have obtained marked or high correlations between different pairs of rankings of their subjects in different stages of practice. Most of the results in this problem have been conveniently tabulated in a recent article by Kincaid[2] and need not be reproduced here. They show convincingly that in various kinds of learning and for both children and adults, and both feeble-minded and superior individuals, correlations between initial and final ability or between ability at any other two measured

[1] Mr. Peterson is wholly responsible for the point of view here taken as well as for the hypotheses and conclusions presented. He has written the article. Mr. Barlow's work in carrying out the experimentation attributed to him in the text was so great that it has been considered fair to make him joint author. Grateful appreciation is here expressed to Dr. S. C. Garrison and to Professor Carl L. Adams.

[2] Kincaid, Margaret, "A study of individual differences in learning." *Psych. Rev.*, 32: 1925, 34-53.

stages are positive and generally rather high. This means that, with the possible exception of persons starting out with very unequal practice, comparatively few individuals change markedly their relative ranks throughout the learning (if, of course, the homogeneity of the group is not too great). The results given agree with data from rather extensive experiments by Pyle, Peterson, and Race, with learning processes ranging from physical performances, like ball-tossing, marble-sorting, and card-sorting, through a variety of activities to the more specifically mental operations of sentence completion, calculations, and answering questions on facts in history.[3] Miss Race concludes her monograph by saying: "In the relation of improvability to initial ability in the function itself, we find that the measure of initial ability is a measure and a prophecy of improvement, that native capacity in any function determines what shall be accomplished by that function" (p. 55)—a conclusion which, though it may hold for the functions investigated, may or may not apply to "any function" whatsoever.

The degree of correlation between rankings of subjects in different stages of practice on a given performance depends, of course, on the complexity of the mental functions involved, on the period elapsing between the two practice stages chosen for correlation, on the relative extents of practice in the particular performance studied at the beginning of the experiment, and on other such conditions. It is obvious, to take an example from the proximity of the two practice stages chosen, that higher correlations would be obtained between the forty-ninth and the fiftieth practices than between the first and the fiftieth or the twentieth and the fortieth, or even between the first and the second, because at the beginning of practice there is more unevenness in exercise on the particular performance than at the final stages of practice.

But these positive, and usually high, correlations between the ranks of the different subjects in various stages of learning do not necessarily show, as some investigators have supposed, that the subjects are becoming more *alike* with practice. They might be either

---

[3] Pyle, W. H., *The Psychology of Learning*, 1921, 184 ff.; Peterson, Joseph, "Experiments in ball-tossing: the significance of learning curves." *J. Exper. Psych.*, 2: 1917, 178-224; and Race, Henrietta V., *Improvability: Its Interrelations and Its Relations to Initial Ability*. Teachers College, Columbia University, Cont. to Educ., 1922, No. 124.

growing more or growing less like one another and still maintain the same rank-order. Of course, some of their learning curves might cross and still permit a fairly high correlation between their ranks. Moreover, the degree of homogeneity of the group of individuals studied will greatly affect the correlations obtained; the crossings in learning curves may be chiefly due to inter-changes of positions of subjects who are natively nearly equal, while subjects differing more innately may not change their relative positions at all. This latter condition seems to obtain in some of our own unpublished results with substitution tests.

If we put the question as to whether practice reduces individual differences in this form: "Do individuals of the highest mental levels in general show the greatest *improvability* under practice?" we get discordant answers from different investigators. This divergence of opinion is partly due to the fact that quite different methods of measuring progress are used—some of them wholly unjustifiable methods. The problem is really not a simple one, and we need not be surprised at the use of different methods of measurement of improvement.

A few examples may reveal the nature of the difficulties involved. Suppose that, in typewriting, Subject A with an Alpha (intelligence) score of 130 improves from 40 to 55 words per minute, while B with an Alpha score of 170 improves from 60 to 80 words per minute. If we also put these scores into another form, certainly as acceptable as words per minute, say number of minutes to write a page of 360 words, we get the data shown in Table I.

TABLE I.—DIFFERENT METHODS OF MEASURING IMPROVEMENT

Subject	Alpha Score	Initial Test Score	Final Test Score	Absolute Gain	Percent Gain
A	130	40 words per min..	55 words per min..	15 words.	37.5
B	170	60 words per min..	80 words per min..	20 words.	33.3
A	130	9.00 min. to page.	6.55 min. to page.	2.45 min.	27.2
B	170	6.00 min. to page.	4.50 min. to page.	1.50 min.	25.0

Which person, according to Table I, makes the greater gain or improvement? What is the percentage of improvement of each subject? It has been the practice of some writers to say in such a case that A makes less absolute gain, and that therefore (A here

being representative of subjects with low intelligence in general, since our present figures are only illustrative) subjects with the most intelligence gain most. Other writers, having their data couched in the second form, have maintained that subjects with least intelligence (as measured by some particular intelligence test) make most progress. Moreover, the percentage of improvement (as the reader will readily see from our figures in the table, and as will become more clear if he imagines slightly different absolute scores and figures) often contradicts the conclusion reached by a study of the absolute 'improvement' or 'gain' (whichever is preferred: they are not the same). And so the confusion grows with the accumulation of new data as long as we are not in agreement on the method to be used.

The confusion resulting from comparing absolute gain in *amount of work done per unit of time* with *amount of time required to do a given piece of work,* I clearly pointed out and illustrated copiously with graphs ten years ago,[4] but the warning then sounded and later emphasized[5] has been heeded by only a few investigators. The confusion still exists in most recent studies. This is due to the fact that the data yielded by any learning process may take either of two forms: (1) average attainment, or accomplishment, per unit of time (or per given number of trials), and (2) time period (or number of errors) for the accomplishment of a given unit of work. Those investigators who put their data into the first form—average attainment—generally find that individual differences increase with practice. This is due to the fact that since the means are very small in the beginning the additions to them (and to the several scores) brought about as the result of practice are at first, absolutely, very small, however large the percentage gain may be. Thus, in ball-tossing, if a group starts with an average of two catches per trial, the absolute gains will, for several trials, be only fractions of one unit, whereas when the average number of catches reaches over 40 per trial, the ensuing gains in successive trials will be about ten units. On the other hand, those who adopt time-units for doing a given amount of work will begin with large units which

---

[4] Peterson, J., *Op. cit.*

[5] Peterson, J., "Thurstone's measures of variability in learning." *Psych. Bull.*, 15: 1918, 452-455; also Peterson, J., "Johnson's measurement of rate of improvement under practice." *Jour. Educ. Psych.*, 15: 1924, 271-275.

show considerable absolute decrease from trial to trial, and these unit changes will regularly decrease in amount until they finally become very small. In a study of ball-tossing the average number of catches per trial was 2.0 for the first practice period of 200 catches and 2.3 for the second period—a gain of only .3; but the average number of errors for the first practice period was 102.5 and 84.5 for the second,— an absolute reduction of 18. The measures of variability—standard deviation or quartile—follow the same law of change, and therefore no measurement of increase or decrease in individual differences with practice can be based on absolute units, however these are expressed.

In view of this fact, a percentage variability measure is to be recommended,—standard deviation divided by the mean, or quartile divided by the median. Although technical objections, which need not receive attention here, may also be given to the use of this method, nevertheless, it seems to be the most serviceable and the least deceiving of methods now in use.[6] It is also advisable to make careful studies of the scores of a few individuals occupying extreme positions in a group, since it is possible for such individuals to diverge with practice, even though the great majority of the members of the group converge and the coefficient of variability gradually decreases. The possibility of this misinterpretation of extreme cases is particularly to be kept in mind when the method of dividing quartiles by medians is used, because by this method the upper and lower twenty-five percents are disregarded. When a sufficiently large number of subjects is used, a statistical check on the possibility of such a misinterpretation is to divide half of the difference between the 90th and the 10th percentiles (or even between the 95th and the 5th) by the median.[7] If this measure of

[6] The method of giving percentage gain from an initial score, although frequently used, is very erroneous and deceiving. When applied to the two kinds of units mentioned above, it gives a different result for each method, for reasons which become obvious from our illustration of A's and B's scores in Table I.

[7] While the semi-interquartile range (Q) divided by the median does not take account of the changes outside the middle fifty percent of the group, as does the coefficient derived from dividing the σ by the mean, a number of calculations by these two methods show that the two measures agree from group to group very closely—closely enough for the rough determinations of the present study. Kelley has shown ("A new measure of dispersion," *Quart. Pub. Am. Stat. Assoc.*, 17: 1921, 743-749) that the most reliable interpercentile range is approximately 93P-7P, and that this is but very slightly inferior in reliability to the more convenient 90P-10P, which we have used

variability increases while the quotient of the quartile over the median decreases, we have a clear indication that the extreme cases are changing somewhat differently from cases near the central tendency.

Miss Kincaid, in the article referred to, has shown that, treated from the standpoint of changes in variability coefficients, most published results on the problem at issue here actually show that individuals become more alike under practice. The behavior of extreme cases was, however, not considered. One of the present writers (Peterson) had much earlier come to the same conclusion with respect to two investigations studied in some detail,[8] and a more recent study,[9] comprising a statistical examination of numerous previous results as well as some original experimentation, had also found, before Miss Kincaid's article was published, that in most investigations yet carried out practice has tended to reduce individual differences.

Summary of Previous Investigations

A brief summary of the results of previous investigations is all that is necessary here. The results given in Table II have in most cases been worked out by us or by Miss Kincaid into variability coefficients from the original data as published by the several authors.[10]

---

to throw light on the variability (divergence or convergence) changes in the near extremes of the distributions. The latter measure is more convenient because of the inclusion of the 10th and 90th percentiles in the usual percentile tables. Our results show that occasionally (90P-10P)/Md may show divergence or convergence not reflected by the Q/Md coefficient, or even a tendency that is the opposite of one indicated by Q/Md. The P.E. of the percentile range between the 10th and the 90th percentile is $.600D/\sqrt{n}$, D standing for the percentile range and n for the number of cases. From the median and the two percentiles in question the skewness may be directly found. There is no skewness if the difference between the average of the 90th and 10th percentiles and the median is approximately zero, or if this difference is small in relation to its P.E., the value of which is $.40412\ D/\sqrt{n}$. Also, for mesokurtic distributions $Q = .26315\ D$. If Q/D is less than .26315, the distribution is peaked, or leptokurtic; if greater than .26315, it is flat-topped, or platykurtic. These specific determinations, however, are of less value to our present problem than the noting of changes in kurtosis with practice. The P.E. of Q/D is $.18736/\sqrt{n}$. (Kelley, *op. cit.*)

[8] See footnote 5.

[9] Master's thesis by M. C. Barlow, *Individual Differences as Affected by Continued Practice*, 1924, in the library of George Peabody College for Teachers.

[10] For formula for P.E. of the coefficient of variability, see Reitz, *Handbook of Mathematical Statistics*, p. 77.

TABLE II.—SUMMARY OF PREVIOUS INVESTIGATIONS OF THE EFFECTS OF PRACTICE

Nature of Practice	Author[8] and Date	Subjects	Unit of Measure	Trials Compared	By Whom Calculated	Variability Coef., σ/m.
1. Subjects become more alike						
Color-naming	Chapman '14	22 adults	Attainment	Av. first and last three	Peterson	.15, .10[1]
	Gates, G. '22	23 adults	Time	First and last	Barlow	.21, .17
Cancellation	Chapman '14	22 adults	Attainment	Av. 1–3, 8–10	Peterson	.13, .11[1]
						.13, .05[1]
	Thorndike '16	11 adults	Attainment	First and last	Kincaid	.13, .10
						.12, .08
	McCall '16	88 children	Attainment	First and last	Kincaid	.51, .21
						.30, .18
Giving opposites	Chapman '14	22 adults	Attainment	Av. first and last three	Peterson	.14, .10[1]
Giving analogies	Perrin '19	22 adults	Time	First and last	Peterson	.48, .32[2]
Mirror-reading	Perrin '19	22 adults	Attainment	First and last	Peterson	.28, .15[2]
	Johnson, O. '23	60 adults			Peterson	Converge
Mental multiplication	Whitley '11	9 adults	Time, errors	First and last	Kincaid	.55, .37
Mental multiplication	Chapman '14	22 adults	Attainment	Av. first and last three	Peterson	.28, .20[1]
Mental multiplication	Gates, G. '22	23 adults	Time	First and last	Barlow	.45, .43
Addition	Gates, G. '22	23 adults	Time	First and last	Barlow	.16, .12
Telegraphy	Thurstone '18	165 adults	Attainment	First and last	Peterson	.74, .52
Typewriting	Thorndike '16	11 adults	Attainment	First and last	Kincaid	.41, .18
	Chapman '19	19 H. S. pupils	Attainment	25th and 175th	Peterson	.16, .07[1]
Card-sorting	Brown, Warner '14	26 adults	Time	First and last	Barlow	.13, .10
Copying addresses	McCall '16	88 children	Attainment	First and last	Kincaid	.36, .23
Maze-learning	Whitley '11	9 adults	Time, errors	First and last	Kincaid	.30, .25
Braille-writing	Kincaid '25	?	?	First and last	Kincaid	.29, .17
Dart-throwing	Kincaid '25	?	?	First and last	Kincaid	.38, .31
2. Nature of change in doubt						
Card-sorting	Myers, G. '18	27 adults	Attainment	First and last	Myers	.12, .12
Javelin-throwing	Johnson, B. '23	14 adults	Attainment	First and last five (av.)	Peterson	.17, .12[2][6]
					Peterson	.16, .19[2][7]
Form-sorting	Woodrow '17	20 (F.M.) children	Attainment	First and last	Kincaid	.14, .18
		16 children	Attainment	First and last	Kincaid	.13, .13

TABLE II.—SUMMARY OF PREVIOUS INVESTIGATIONS OF THE EFFECTS OF PRACTICE—*Continued*

Nature of Practice	Author[8] and Date	Subjects	Unit of Measure	Trials Compared	By Whom Calculated	Variability Coef., σ/m.
3. Subjects become more unlike						
Addition	Thorndike ... '10	19 adults	Attainment	First and last	Kincaid	.26, .29
Addition	Thorndike and Donivan ... '13	21 boys	Attainment	Av. 1–2, 31–32	Peterson	.30, .32[2]
Addition	Chapman ... '14	22 adults	Attainment	Av. first and last three	Peterson	.19, .20[1]
Addition	McCall ... '16	88 children	Attainment	First and last	Kincaid	.45, .51
Addition	Race ... '22	43 children	Attainment	First and last	Barlow	.52, .62
Mental multiplication	Thorndike ... '08	28 adults	Time	First and last	Barlow	.34, .47
Multiplication by substitution	Race ... '22	43 children	Attainment	First and last	Barlow	.17, .35
Javelin-throwing	Murphy ... '16	10 adults[3]	Attainment	Av. 1–50, 51–100, 101–150	Peterson	.59, .63, .66
		10 adults[4]	Attainment	...	Peterson	.55, .54, .63
		10 adults[5]	Attainment	...	Peterson	.57, .59
Ball-tossing	Peterson ... '17	28 adults	Attainment	Av. 1–5, 26–30, Practice periods	Peterson	.54, 1.15 .48, 1.07[2]

(1) V = Q/Md.; (2) V = M.V./M.; (3) Subjects practiced five times per week; (4) Subjects practiced three times per week; (5) Subjects practiced once per week; (6) Right hand; (7) Left hand.

(8) References to the authors cited are as follows:

Barlow, M. C. (Master's thesis, Peabody College Library, 1924).
Brown, W., *Univ. of California Publ. Psych*, 1914, No. 4, 269-321.
Chapman, J. C. *Teachers Coll. Contr. to Educ.*, 1914, No. 63; also ***Jour. Appl. Psych.*, 3:1919, 252-268.**
Gates, G. S., *Arch. of Psych.*, 1922, No. 58.
Johnson, Buford, *Psych. Rev.*, 26: 1919, 300-316.
Johnson, O. J., *Jour. Educ. Psych.*, 14: 1923, 540-554.
Kincaid, Margaret, *Psych. Rev.*, 32: 1925, 34-53.
McCall, W. A., *Teachers Coll. Contr. of Educ.*, 1916, No. **79.**
Myers, G. C., *Amer. Jour. Psych.*, 29: 1918, 316-326.
Murphy, H. H., *Jour. Educ. Psych.*, 7: 1916, 150-162.
Perrin, F. A. C., *Psych. Rev.*, 26: 1919, 42-62.
Peterson, J., *Jour. Exp. Psych.*, 2: 1917, 178-224; also *Psych. Bull.*, **15: 1918, 452-455,** and ***Jour. Educ. Psych.*, 15: 1924, 271-275.**
Race, Henrietta V., *Teachers Coll. Contr. to Educ.*, 1922, No. 124.
Thorndike, E. L., *Amer. Jour. Psych.*, 19: 1908, 374-384; *ibid.*, 21: 1910, 483-486, and ***ibid.*, 27: 1916, 550-565.**
Thorndike, E. L. and Donovan, M. E., *Amer. Jour. Psych.*, 24: 1913, 426-428.
Thurstone, L. L., *Psych. Bull.*, 15: 1918, 210-212.
Whitley, M. T., *Arch. of Psych.*, 1911, No. 19.
Woodrow, H., *Jour. Educ. Psych.*, 8: 1917, 85-96, 151-165.

The quantitative results in terms of variability coefficients in the table would vary slightly when calculated by different methods. For instance, the coefficients of an average score of the first five trials, say, as compared with those of the last five would give slightly different (and more reliable) results than would coefficients of the initial and final single trials; moreover the variability measures obtained by dividing the mean variation or the standard deviation by the mean (m.v. or σ/m) will be somewhat different from those resulting from dividing the quartile by the median (Q/md) since the former are determined by the scores of all the individuals while the latter are directly influenced by the scores of only a few individuals and are not a function of extreme variations. The percentage variability is also influenced by the stages of the practice at which measurements are taken. The degree of homogeneity of the subjects' abilities, ages of subjects, and many other factors also have their effects both on the size and the changes of variability coefficients, and must ultimately have much more careful attention than they have thus far received.

Investigations by the Writers

1. A Preliminary Experiment

Some time ago we began at Peabody College some preliminary investigations of the influence of practice on individual differences. Having reason to believe that the practice effect would differ with the degree of complexity, we selected for practice a number of different performances which were supposed to represent different degrees of complexity. The experiments were all carried out under conditions so well controlled that the experimenter (Mr. Barlow) could assure himself that each subject worked at approximately his maximal ability. They concerned practice in (1) digit-symbol substitution (material of Pintner and others), (2) card-sorting, and (3) mental multiplication. In the first exercise each subject made 200 substitutions in each of six consecutive days, and marked his position at the end of each minute, so that his speed could be determined. In the second exercise each subject distributed 120 cards daily for six days. The cards were distributed into an upright box 30 by 48 inches and 6 inches deep, divided into forty compartments—five rows of eight in each row, each compartment

being approximately five inches square. The rows were called A, B, C, D, and E, in order, and the holes in each one were numbered from 1 to 8, but no labels were put on the box; the subject had to discover and then remember the letter and number of each hole. To make the problem clear, each subject was given a fore-exercise consisting of the distribution of five cards. In the third exercise, the multiplication consisted of computing mentally the products of two-digit numbers by two-digit numbers. Each subject completed twelve sheets of 40 examples each, working at periods of twenty minutes daily till the work was completed. The periods ranged from twelve to twenty. The score was the number of seconds required to work forty problems, as determined from the twenty-minute records of problems and part-problems completed. The percentage of errors in multiplication and in substitution was tabulated, but did not figure in the score. In the card-sorting the errors were noted, but not corrected. The subjects were first given a little practice in mental multiplication, after seeing the method illustrated by the experimenter before the group; but each subject was left free to use the method most easy to himself. The card-sorting exercises were conducted individually; the other exercises, in groups.

The subjects in these experiments were 20 twelve-year-old children in the sixth-grade of the Peabody Demonstration School. Their intelligence had been determined by the Stanford Binet and three different forms of the National Intelligence Test. The mental ages of the children as determined by the four intelligence tests, each given equal weight, were: median, 154 months; Q_1, 148.3 months; Q_3, 177.5 months.

The average of all the correlations between the different pairs of the intelligence tests was .61, while that of the correlations between successive performances (on different days) in the practice exercises were as follows: .87 for digit-symbol, .87 for card-sorting, and .86 for mental multiplication. Correlations of initial and final practice results are interesting. Initial digit-symbol and card-sorting correlation coefficient is .79; the final correlation, .24. Initial digit-symbol and mental multiplication is .12; the final, .16. Initial card-sorting and mental multiplication is −.04; the final, .67.

Now we come to the significant question: Do these children become more alike or more unlike with practice in the functions

involved in our practice exercises? The coefficients of variability, σ/m, for six successive days of 200 substitutions each are: .16, .14, .19, .17, .18, and .18; average of first half, .16, of last half, .18. The six days of card-sorting for the 20 subjects gave coefficients of variability in order as follows: .19, .14, .11, .11, .13, and .12; average of first three, .15; of last three, .12. The mental multiplication exercises gave coefficients for successive groups of 80 problems of .26, .30, .30, .31, .33, and .36; average of first three, .29; of last three, .33. Our results, then, show a divergence in the individual abilities with practice in substitution and in mental multiplication, but a convergence in card-sorting. The coefficients for successive sixths of the entire practice, however, are consistent only in the case of mental multiplication.

2. Another Experiment

Our next results are from a more extensive experiment in mental multiplication on 101 subjects, but owing to irregularity in the attendance of five, we include the scores of only 96 in this report. These 96 students had a median age of 175.9 months; Q_1 and Q_3 were 166.5 and 182.5 months, respectively. They were distributed in grades as follows: 7th, 10; 8th, 37; 9th, 23; 10th, 23; 11th, 2; and 12th, 1. The practice experiments were administered by one of the writers (Barlow) and an assistant, under practically the same conditions described above for mental multiplication, except that in this case 63 examples, instead of 40, were given on each sheet, and the time recorded was that needed by each subject to complete a sheet. The rapid students were allowed to work more groups of problems than did the slow ones. The results (Table III) have therefore been tabulated in four groups for careful study by one of the writers (Peterson). The groups are as follows: A, the 42 pupils who worked through 13 units of 126 problems each; B, the 24 pupils who completed 10 to 12 units; C, the 30 pupils who completed fewer than ten, but as many as 7 units. C is, of course, the slowest group. Finally we give in Group D the data for all 96 cases through the first 7 units of work covered by all the subjects. Coefficients of variability are figured both from the quartile and from the 90- and the 10-percentile scores, since the former method shows the variability of only the middle fifty percent of the group.

TABLE III.—MEDIANS AND COEFFICIENTS OF VARIABILITY THROUGH SUCCESSIVE UNITS OF WORK IN MENTAL MULTIPLICATION BY 96 SCHOOL CHILDREN OF BURLEY, IDAHO

	Minutes for Successive Work-Units of 126 Problems in Each Unit														
	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean	
Group A, No.	42	42	42	42	42	42	42	42	42	42	42	42	42	(1–6)	(8–13)
Median*	46.5	37.0	33.7	32.6	32.3	29.7	29.7	29.3	28.0	28.0	27.0	24.7	24.1	......	......
Q/Md.	.19	.18	.16	.10	.13	.11	.10	.11	.11	.09	.09	.11	.08	.145	.098
90P–10P/2Md	.28	.28	.25	.20	.22	.20	.21	.21	.21	.20	.21	.21	.19	.238	.205
Group B, No.	24	24	24	24	24	24	24	24	24	24	......	......	......	(1–5)	(6–10)
Median	58.0	49.0	46.5	43.0	39.0	37.0	36.2	34.0	35.0	36.0	......	......	......	......	......
Q/Md.	.12	.08	.10	.07	.09	.12	.13	.11	.13	.10	......	......	......	.092	.118
90P–100/2Md	.20	.21	.16	.14	.15	.22	.18	.16	.21	.21	......	......	......	.172	.196
Group C, No.	30	30	30	30	30	30	30	......	......	......	......	......	......	(1–3)	(5–7)
Median	72.0	58.7	55.0	49.0	52.0	47.0	46.0	......	......	......	......	......	......	......	......
Q/Md.	.16	.14	.09	.10	.14	.10	.15	......	......	......	......	......	......	.139	.130
90P–10P/2Md	.23	.24	.20	.21	.22	.21	.27	......	......	......	......	......	......	.223	.233
Group D, No.	96	96	96	96	96	96	96	......	......	......	......	......	......	(1–3)	(5–7)
Median	55.5	46.7	43.0	40.3	37.5	35.0	35.0	......	......	......	......	......	......	......	......
Q/Md.	.18	.21	.20	.17	.13	.19	.18	......	......	......	......	......	......	.197	.167
90P–10P/2Md	.37	.38	.33	.29	.38	.33	.32	......	......	......	......	......	......	.360	.343

*The coefficient of correlation between the medians and the quartiles is .48. The corresponding *r's* for Groups B, C, and D are, respectively, .23, .78 and .89.

The ratio of the average of the first six coefficients of Group A to that of the last six is .68, showing a convergence of the middle fifty percent with practice. The ratio of the corresponding averages of the coefficients derived from the 90th and 10th percentile, $\frac{90P - 10P}{2M}$, is .86, revealing a less marked convergence among the extreme half of the group. This fact is perhaps partly explained when one notes that there is actually a divergence with practice in the individuals of our Group B, of lesser ability in mental multiplication; the corresponding ratios in this group are 1.28 and 1.14. For the more able students the process of mental multiplication is simple enough for the members to become more alike with practice, whereas for the less able group it is too complex for the exercise thus to overcome individual differences. But we cannot be sure of this interpretation, for Group C shows no appreciable change in variability one way or the other. The large group, D. however, shows, in conformity with results of most previous investigations, that our group on the whole becomes more alike in mental multiplication with practice.

Our own experiments thus far presented are of the most valuable sort because the measurements were derived directly from the practices themselves. If one uses any extraneous test of the practice effect, one is always liable to the error of using tests at the different stages of training which are not fully and equally representative of the changing functions studied, and also to the error of employing tests not equally fair in degree of difficulty at the different stages of practice. Such tests will not give equally representative scatter of the subjects at the different stages of training. For instance, either a too easy test or a test that does not give psychologically equal increments in degree of difficulty to successively higher percentiles of the group may 'bunch' the subjects relatively too much at certain stages, even though the subjects actually do not converge or diverge with practice. Even the use of the same test-scale for the different stages of the practice is not a guarantee against the possibity of the sort of errors mentioned. However, we shall now present results of practice tested in some cases by standardized tests and in others by the conventional content examinations.

3. Results from Achievement Tests

Four successive groups of children have passed through Grades V to VIII, inclusive, in Peabody Demonstration School, taking precisely the same achievement tests since the beginning of the fifth grade in 1919. The only difference in the curriculum of these four successive groups through the four grades mentioned was the addition of two hours a week of home science to the eighth-grade programs of the last two groups. The following achievement tests were given to each group in every grade at as nearly as possible the same time of year (January): The Ayres Spelling Scale (twenty words selected from the column marked "73%" for the grade), The Nassau County Supplement to the Hillegas Composition Scale, the Woody-McCall Mixed Fundamentals arithmetic test, and the Monroe Reasoning Test in Arithmetic. These several tests, as well as intelligence tests, were administered either by, or immediately under the direction of, Dr. S. C.

TABLE IV.—MEAN SCORES AND VARIABILITY COEFFICIENTS OF 94 CHILDREN THROUGH GRADES V TO VIII

Test		Grades				Interpretation
		5	6	7	8	
Spelling (Ayres)	Mean	15.8	15.7	14.7	15.5	
	M. V./M	.17	.19	.22	.17	No change
	$\frac{90P-10P}{2M}$	.25	.32	.37	.25	
Composition (Hillegas)	Mean	4.0	4.7	5.0	5.8	
	M. V./M	.16	.18	.15	.14	Convergence
	$\frac{90P-10P}{2M}$	.25	.23	.22	.23	
Mixed Fundamentals (Woody McCall)	Mean	22.3	25.9	29.1	29.0	
	M. V./M	.17	.13	.08	.08	Convergence
	$\frac{90P-10P}{2M}$	.27	.19	.12	.14	
Arithmetic Reasoning—Accuracy (Monroe)	Mean	15.1	12.5	16.8	8.9	
	M. V./M	.26	.27	.23	.29	No change
	$\frac{90P-10P}{2M}$	.40	.40	.39	.51	
Arithmetic Reasoning—Principle (Monroe)	Mean	24.5	18.1	22.6	18.5	
	M. V./M	.25	.21	.19	.23	Convergence (slight)
	$\frac{90P-10P}{2M}$	.43	.33	.29	.38	

Garrison. The responses were scored by students under his supervision. Scoring and tabulation were carefully checked and rechecked to eliminate errors. We are indebted to Dr. Garrison for the use of these data, and immediately to Professor Carl L. Adams of East Carolina Teachers College for furnishing us with the means and mean deviations of these tests, from which we have calculated the coefficients of variability. The children were above normal in intelligence, averaging slightly over 107 I.Q., with a first and ninth decile of 92 and 132 I.Q., respectively. In all, 110 children were included in the four groups, but we have complete data for only 94, distributed in the chronological order of the groups according to sex as follows: girls, 15, 18, 10, 7, total, 50; boys, 7, 10, 13, 14, total, 44. It is to be noted that the girls decrease (with one exception), and the boys increase, through successive groups. Table IV gives the relevant data for our present purpose.

TABLE V.—MEDIANS AND COEFFICIENTS OF VARIABILITY OF SCORES IN SUCCESSIVE COMPETITIVE EXAMINATIONS AND IN CERTAIN INTELLIGENCE TESTS BY 224 STUDENTS IN GENERAL PSYCHOLOGY

Section	Examination											Tests		
	1	2	3	4	5	6	7	8	9	10	Average	Alpha	Otis	Atkinson
	Number of Cases													
A	55	58	58	55	61	61	61	60	60	62	62	62	......	......
B	52	52	52	51	50	50	49	49	51	....	53	53	49	53
C	46	47	46	48	49	41	46	49	....	....	49	49	......	......
D	49	45	50	51	51	49	50	51	52	53	60	60	50	......
	Median Scores*													
A	17.3	12.9	12.9	11.3	13.3	18.2	14.1	11.5	18.1	14.2	14.2	125.0	......	......
B	14.8	19.0	16.8	16.7	17.4	19.8	16.5	16.3	19.2	....	17.5	129.4	54.5	13.3
C	17.0	11.8	19.4	10.8	18.2	11.8	13.0	18.1	....	....	14.7	141.7	......	......
D	13.4	13.8	12.0	12.5	14.3	21.3	19.6	16.7	16.0	14.3	14.4	130.0	58.5	......
	Coefficients of Variability (Q/Md)													
A	.29	.22	.21	.39	.22	.14	.19	.24	.14	.17	.13	.14	......	......
B	.35	.32	.12	.22	.18	.21	.28	.19	.11	....	.13	.12	.09	.19
C	.23	.34	.19	.26	.10	.29	.20	.11	....	....	.11	.12	......	......
D	.33	.33	.28	.22	.25	.28	.19	.20	.27	.21	.19	.12	.07	......

*Correlations between medians and quartiles, A, $-.52$; B, $-.20$; C, $-.24$; D, $.62$

Table V presents the number of cases, the medians, and the variability coefficients of a series of successive competitive examinations in general psychology and of certain intelligence tests and

one (Atkinson) ingenuity test. The results from four sections in psychology—A, fall of 1923; B, fall of 1924; C, spring of 1925; and D, spring of 1926—are given, involving the records of 224 students in Peabody College, mostly sophomores. The sections were all taught by one of the writers (Peterson), and the examinations, each with a maximum of 30 points, were prepared by him and scored by an assistant. The papers were regularly returned at the next recitation following the examination, ranked in the order of the number of points made. By this means a healthy competition is kept up throughout the course.

In the grade subjects tested by standard tests (Table IV) there are two cases, spelling and arithmetic reasoning, in which accuracy is counted, that show no change in variability through the four years, but the other three show convergence. There is most clear and unambiguous convergence in the case of the Mixed Fundamentals test in arithmetic. The results of such tests running through years must be interpreted in the light of another fact: that is, that in intelligence-test scores there is probably a rather marked convergence with increasing age in childhood. This tendency is noticeable in the data from Myer's *Measuring Minds,* 1921, page 27, from which we have calculated means, standard deviations, and coefficients of variability:

Age in Years	6	7	8	9	10	11	12	13	14
Number of cases	516	1191	1237	1440	1304	1318	1169	1204	928
Means*	11.5	17.0	23.9	29.6	34.8	38.7	44.5	47.8	50.8
σ	8.1	9.7	11.6	12.1	13.4	13.2	14.4	15.0	14.9
σ/m	.70	.57	.49	.41	.39	.34	.32	.31	.92

*Correlation between the standard deviations and means is .97.

If the Myers test, which is the same for all ages, is to be relied upon, we must make allowance in any performance tests for this reduction in variability with the advance of age in growing years, and it is then possible that the increasing likeness in the less marked cases of convergence in Table IV is wholly due to this fact of convergence in mental ability independently of the effects of training, or even in spite of it. Myers tests, however, may not present enough room at the top of the scale to give opportunity for the older children to represent their abilities as well as the younger children.

In the case of the psychology students, all adults, there is a clear convergence in variability of scores with progress through the twelve weeks' course; that is, the students under training were becoming more alike in their abilities in psychology. Even though the variabilities which we calculated from the first and the ninth decile show this same tendency (we have omitted these data as they add nothing to those given), we cannot conclude that a few extreme individuals do not show an opposite tendency. It will be noted that the variability coefficients of the average scores in psychology are about equal to those in Army Alpha, but that the Otis Test scores have very much less variability than these. This may be due to the greater homogeneity in method or to possible differences in materials employed in the test, in appropriateness of time allowed, in degree of equality of units employed, or what not. At any rate, the difference in variability coefficients is not only marked, but consistent, so far as our data go; and it points plainly to a possible danger in using extraneous tests of any kind, even conventional examinations (for there is considerable variability in the coefficients of the several psychology examinations), to measure effects of practice on individual differences. If the successive tests do not give similar variability coefficients, they may lead to serious misinterpretations, and if they do exhibit gradual changes one way or another we cannot be sure this is not due to the nature of the extraneous tests, rather than to actual change in the variability of the subjects in the function studied. The psychology tests did not, of course, cover all the content of the course, neither were they uniform in difficulty as the medians and the correlations between medians and quartiles show. It is impossible, without careful standardization, to make examination questions equal. Moreover, tests of various kinds are apt to bunch or distribute subjects differently according to different degree of ability, either within certain ranges only or within the total range. These shortcomings, varying in different examinations and standardized tests, produce a fluctuation in the coefficients of variability at different stages that does not in any true sense reflect changes produced by practice only.[11]

[11] The low, and even negative, correlations between the quartile deviations and the medians in the psychology examinations show clearly the inadequacy of such examinations as determiners of changes of variability in learning. Besides being what we have called "extraneous" measures, these examinations have another defect. In the first part of the course they were technical and

They must therefore be scrupulously avoided by means of adequate experimental control, rather than by compensatory elaborations in statistical methods of handling the data, if investigators are to agree on the meaning of the results. Direct measurement of the performances under study themselves, at different stages or continuously, is to be preferred to the use of what we have called "extraneous" tests, however well these latter are standardized.

Comments and Conclusions

Our data have disappointed us by yielding no clean-cut results. Indications as to improvement of technique are, however, preferable to overhasty generalizations. Research suffers more from premature conclusions than from indefiniteness in results and consequent suspension of final judgments. Our results do serve to emphasize the necessity, in future researches along this line, of obtaining scores by the several subjects at the different stages of the practice in terms of the whole unit of work actually done, not by mere samplings by means of extraneous tests of any kind. On the basis of this caution we have grave doubts concerning the validity of such conclusions as Kelley's,[12] for instance, that nurture has reduced native differences in computation functions more than it has the native differences in such functions as are involved in history and language information.

We began investigating our present problem a few years ago with the conviction (resulting from certain preliminary data at hand) that subjects of normal heterogeneity would become more alike with practice on the simpler processes or activities, but more

---

quantitative, while in the latter part of the course they gradually changed toward the qualitative type, calling for criticisms, evaluations, explanations, etc. Thus, in the later examinations discrimination between very good (or very poor) answers and those at the central tendency was more difficult. This results, as numerous studies have shown, in two things: the raising of the general average score, on the one hand, and a lessening of differences, on the other. The decrease in the variability coefficient may be only a reflection of these changing aspects of the examinations. The low correlation between quartiles and medians is easily accounted for in the lack of uniformity in the successive examinations with respect to their quantitative aspects. Several examples studied show that, as a rule, a hard and highly quantitative examination brought the average down considerably and at the same time increased the quartile deviation.

[12] Kelley, T. L., *The Influence of Nurture upon Native Differences*, 1926, 23 ff.

different on the more complex activities. We naturally supposed, then, that if all vocational activities were arranged in a series from the simplest to the most complex, there would be a point in the series for a given individual, with given abilities, below which we could advise him to choose his line of vocational activities if he would hold his own in competition with others. Industrious practice in an activity below this point would be rewarded by his becoming more and more like the most practiced. With certain modifications, particularly with respect to special aptitudes, we still hold this opinion and hence regard the whole matter as of fundamental importance to education. But our own results and those of other investigators have surprised us, and, let us confess it, disappointed us. We must question the correctness of our over-simple original hypothesis. The problem is more complex, and possibly also more important, than we had anticipated. There is a probability that our supposed point in the scale separating the complex from the simple processes for the representative individual was wrongly estimated, was placed too low—a probability that most of the processes we regarded as "complex" (in the sense mentioned) are really "simple" (in the sense that normal individuals converge with practice on them). In addition to this, it is possible also that a greater number of learning processes can be carried on simultaneously by subjects of high, than of low ability, even in cases where practice on single activities results in convergence; or, to put it differently, that even on a *group* of "simple" activities practiced together, normal persons will diverge with practice. General observation supports this view.

It is interesting to note that physical activities, like ball-tossing and javelin-throwing, show divergence with practice. Is this because of innate peculiarities or because in these activities less definite aid can be given the several subjects in their special difficulties, so that each must work out his own salvation? Is it possible, moreover, that one of the main differences in efficiency is the quality of keeping up vigorous activity without constant stimulation? Are the poor learners, even though they may gain on their superior competitors with proper motivation, as a rule inclined to relax effort and to allow themselves to drift into an easy-going pace (whatever that may mean in neurological terminology)? But why

have subjects under practice generally become more alike in mental multiplication and more unlike in addition? Is this because the latter is more complex (an incredible assumption) or because the former is so complex as to demand of the individual so much ingenuity at first as to produce high variability coefficients for later comparisons? It is interesting to note that the coefficients where the subjects diverge with practice are usually large, but there are some conspicuous exceptions. These and many other questions we leave to further research in this interesting field.

In addition to the need of certain precautions already stressed, we desire to emphasize here, for future investigations of the effects of practice, the necessity of studying carefully the correlations between the factors represented in the numerator and the denominator of the coefficient of variability. It is desirable to investigate experimentally the conditions which bring about low or negative correlations, and especially to remove such conditions by improved experimental control in as far as this is possible rather than by compensatory statistical calculations. A special study needs to be made of the relationships between the σ (or D) and the mean (or the median) of time-for-unit-of-work data in the early stages of learning when the time curves are dropping rapidly. Data of the sort needed for the analyses here suggested were in most cases not available in the present work. This deficiency was due in part to our own ignorance as to the need of certain kinds of data at the time the work was begun. Alas, the finer elements of technique in any important piece of work come only with practice in that work, not before it is begun!

PREFATORY NOTE TO CHAPTER XV

This chapter presents a brief survey of the literature (or of some portions of it) that deals with the factors underlying the appearance of one faily well-defined specific trait—musical ability. Dr. Farnsworth concludes that the question is "still a moot one."

The main lines of attack upon this problem have been (1) to investigate family resemblance in musicality, and (2) to investigate the so-called 'elemental character' of musical abilities.

The results from the first approach, while fraught with suggestive implications, have the same inherent limitations that are contained in a similar approach to the heredity of intelligence or of any other mental trait. Nature and nurture simply cannot be extricated from each other by ordinary means of investigation.

The results from the second approach, as the author has pointed out, are also inconclusive, partly because of ambiguity in the minds of investigators as to what actually constitutes training that can test the elemental character of musical abilities.

CHAPTER XV

THE EFFECTS OF NATURE AND NURTURE ON MUSICALITY

PAUL R. FARNSWORTH
Stanford University, Palo Alto, California

There are at least two conflicting views held at the present time concerning the relative effects of nature and nurture on musicality. One, with Watson (38) as its chief exponent, holds that nurture (environment) is all important. "Our conclusion, then, is that we have no real evidence of the inheritance of traits. . . . . I should like to go one step further now and say, 'Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select—doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors.' "

Certain other psychologists, while doubtless willing to give nurture a place as one of the causes of musicality, believe that nature is more potent. In pursuance of this view they have developed 'capacity' tests with which to select children blessed by nature with superior talents.

C. E. Seashore in his *The Psychology of Musical Talent* (27) and in numerous articles states again and again his belief in the heredity of musicality. To quote one passage: "Musical talent is a gift bestowed very unequally upon individuals. Not only is the gift of music itself inborn, but it is inborn in specific types. These types can be detected early in life, before time for beginning serious musical education. This fact presents an opportunity and places a great responsibility for the systematic inventory of the presence or absence of musical talent."

In this chapter I shall present a fairly comprehensive survey of the available data relating to this problem. The material will be grouped for the sake of convenience into a number of divisions.

I. Genetics

Hazel M. Stanton (31) aided by an appropriation granted by the Carnegie Institution of Washington, studied the abilities of eighty-five parents and offspring on four of the Seashore phonograph tests—pitch, intensity, time, and tonal memory. "These measures were supplemented by a systematic interrogation which covered questions in regard to musical environment, musical education and training, musical activity, musical appreciation, musical memory and imagination." Her results are stated as follows: "The harmony of the results with certain Mendelian laws in the family distribution tables of assumed gametic formulae is not improbable."

J. A. Mjöen (21), a firm believer in the biological inheritance of musicality, has worked through many family lines. Among his reports is the rather interesting case of an extremely musical girl in an exceptionally unmusical family. Her case was a mystery until it was disclosed that she was the illegitimate daughter of a very prominent musician.

Valentine Haecker and Th. Ziehen (14), in two extensive surveys, found by the questionnaire method that the numerous musical factors which they analyzed apparently followed Mendelian principles. This work was later extended by Koch and Mjöen (19) with similar results.

Galton (10) believed in the inheritance of musicality, but was handicapped in the investigation of the distribution of musical taste in families because of the numerous and petty jealousies among musicians. Nowhere could he get even a relatively unbiased rating of musicians' abilities. That standards of musical taste are even yet quite relative is well shown by Esther A. Gaw's (11) survey of the reports of New York City music critics on identical concerts.

According to C. C. Hurst (17) "That this natural disposition or temperament [for music] is innate and hereditary, there can be little doubt. Musical associations and careful training can, of course, do much in enriching the musical qualities in an individual of a musical disposition, but in the absence of the musical temperament these outer stimuli are practically powerless." For Hurst, musicality is probably a recessive trait.

William Wallace (37) actively combats the conclusions of Hurst and shows that it is at least as reasonable to postulate nurture as the cause of persistent musicality in family lines as nature. He attributes musicality to the "continuity of vocation."

A similar view is taken by Mrs. E. F. Copp (5). "I think it [musicality] is a matter of heredity, but that almost everyone possesses the heredity." She finds that positive [absolute] pitch can be acquired by 80 percent of normal children. There is a difference, however, in the speed of learning. Individuals lose the 'inborn' talent through lack of training. She points out that many musical people, like Haydn and Schubert, came from unmusical families. The paucity of musical women is regarded as evidence against the argument of musical inheritance.

According to Vaerting (34), lack of creative musical talent in women is correlated with the primitive sexual function of vocal music. The male is the natural musical performer.

Both Hans Rupp (24) and Davenport (7) regard musical talent as inherited. Drinkwater (9) decides that the 'musical sense' is probably recessive.

Stern (32) and Bernfeld (3), however, affirm that musicality may be acquired. The latter states that by the exertion of the will the unmusical may become musical. He cites a number of cases in support of his contention.

Some work of F. O. Smith (30) may be construed at first glance as favoring the environmentalists. Tests of pitch discrimination were given two large groups of siblings. One group gave a correlation of .48, P. E. .031; the other .43, P. E. .035. In each case a younger child was paired with an older sib. When the younger children of group two were paired with other children of the same age and sex as their sibs, but who were not related to them, the correlation was .53, P. E. .030. This surprising result may possibly be explained in part by probable errors in selection. It is quite likely that, on the average, the younger the child, the younger the sib with whom he was paired. Since age is an important factor in test scores of this type (see the Seashore test age-norms), this would in itself give spurious positive correlations. The reliability of his test is also unknown.

Voss (36) cites the case of a musical man who had two wives, one musical and one unmusical. There were seventeen children involved in the five succeeding generations. As in the well-known Kallikak study of Goddard, the effects of social heredity were not separated from biological inheritance; so the scientific value of this study is questionable.

II. Elemental Character by Test

The difficult question of nature and nurture may be examined from the test standpoint, *i.e.*, we may ask when does a test disclose an elemental character? According to Seashore and Mount (28) "to the extent that the test is elemental, there should be no appreciable improvement with practice, development with age or variation with general intelligence."

1. General Intelligence

Of the three criteria, that of general intelligence is perhaps the easiest to meet. If the Stanford Binet be arbitrarily designated as its measure, many tests have been developed which have a low correlation with it.

The work of Mrs. Leta Hollingworth (15) has shown that gifted children (those with high intelligence quotients), as a group, do no better on the Seashore 'capacity' tests (excepting possibly the test of time discrimination) than do children of average intelligence. While one should hesitate to conclude that children of high intelligence are no more likely to be musical than children of average intelligence (Dr. Hollingworth does not make this statement), this investigation at least suggests that conclusion.

In his *Psychology of Musical Talent*, C. E. Seashore reports that not one of his battery of tests correlates to any great extent with estimated intelligence. F. O. Smith (30), however, finds correlations of .70, P.E. .023, between pitch discrimination and general intelligence in the case of boys, and of .63, P. E. .026, in the case of girls. 'General intelligence' was here ascertained by teachers' ratings of reliability and brightness (quickness and accuracy of mental grasp—general wide-awakeness). As teachers' ratings are notoriously unreliable, too much stress should not be laid on this work.

Weaver (39) from a study of 94 subjects, found the following correlations between scores on the Seashore battery of music tests and the Army Alpha: pitch .35, intensity .24, time .12, consonance .06, and tonal memory .26.

The writer has recently correlated scores on the Seashore and Kwalwasser tests with Thorndike intelligence scores. For approximately 150 university-students the results were: pitch, .14; rhythm, .17; memory, .11; consonance —.38 (an elimination of four extreme cases changes this correlation to +.02); time .10; intensity, .11; melody —.08; and harmony .22.

Serial reaction time seems to have musical significance. It may be simply defined as a speed test in which each response becomes the stimulus for the next response in an extended chain of stimuli and responses. C. F. Hansen (13) reports three groups of correlations with this test and the Army Alpha Intelligence Examination. Forty-one 'musical' college students furnished a correlation of .51, P. E. .08; sixty-one high-school stenographic and teachers' college students, a correlation of .20, P. E. .08; and 174 high-school stenographic students, one of .14, P.E. .05. At first glance, one might assume that the largest group gave the most reliable data, and so state that the Army Alpha Test had little correlation with serial reaction time. Another hypothesis might be that the Alpha was a content test for the stenographers, but a speed test for the more advanced students.

This latter hypothesis is somewhat substantiated by work by R. Seashore, M. Tinker, and the writer. With groups of university students ranging in number from 28 to 35, scores on serial action correlated with Part 5 of the Ohio State University Psychological Test, Form 10, .14; with Thorndike, .20; with Otis Group Intelligence Scale, Advanced Examination, Form A, .30; and with the Army Alpha, Form 8, .53.

In some unpublished work, R. Seashore finds a correlation of —.21, P. E. .11, between simple auditory reaction time and Thorndike intelligence scores. From a study of 47 men at the University of Iowa the same worker (29) reports a correlation of .10, P. E. .09, between percentile rank in the University of Iowa Qualifying Examination and rank in 'sense of rhythm' (Seashore test).

Schüssler (25) some years ago started a controversy with Bartsch on the relation between scientific and musical abilities. Ac-

cording to the former's study of 200 school children, the musical are considerably superior to the non-musical in scientific ability. Bartsch (2), from a study of 200 normal-school students who had received six years of musical training, reports that the more musical have no greater scientific ability than the less musical. Schüssler (26) answers Bartsch in a later article.

The study of Cox (6) on the youthful achievement of men deemed great in later life includes a number of musicians. Révész (22) and numerous others write of musical prodigies who seem to have high scholastic intelligence. On the other hand, there are numerous reports of remarkable musical ability accompanied by inferior scholastic intelligence. For example, S. J. Holmes (16) claims that "Blind Tom, who possessed a phenomenal aptitude for playing any piece of music he may have heard, was practically an imbecile."

2. Age

The age factor is not easy to control. Completely to rule it out, would mean the construction of a test which gave the same average score in large groups regardless of the ages of the group members. Even Seashore, in his well standardized tests, has three distinct norms,—for fifth-grade pupils, eighth-grade pupils, and adults (see Table I). He justifies this by the introduction of two concepts,—the *physiological* and the *cognitive* limits. All tests aim to disclose the absolute, or physiological, limits in an ability. They fail to the extent that they become monotonous and uninteresting. What they test then is the cognitive limit. Thus, there is more chance of tapping the adults' physiological limits than those of eighth-grade pupils, and more likelihood of obtaining the physiological limits of eighth than of fifth-grade pupils.

TABLE I.—TYPICAL PERCENTILES IN THE SIX SEASHORE TESTS FOR FIFTH-GRADE PUPILS, EIGHTH-GRADE PUPILS, AND ADULTS

Percent Responses Correct	Pitch			Intensity			Time			Rhythm			Consonance			Memory		
	5	8	Ad.	5	8	Ad.	5	8	Ad.	5	8	Ad.	5	8	Ad.	5	8	Ad.
55	21	7	3	8	...	....	17	...	....	11	4	....	27	11	4	59	27	14
70	59	31	15	38	8	3	75	45	13	76	50	23	83	66	56	82	64	41
80	87	63	45	73	36	14	97	90	61	97	88	77	98	97	97	94	84	65
95	100	100	100	100	98	92	100	100	100	100	100	100	100	100	100	100	99	96

The statement by F. O. Smith (30) that "the amount in favor of the adult shown in all group statistics is amply accounted for by the difficulty in making a reliable test on the young and by their lack of information" seems hard to reconcile with the fact that in intelligence tests the average mental age of the eighth-grade pupil is found to be almost, if not quite, that of an adult.

The work of Hollingworth, reported earlier in this chapter, also tells against Smith's assumption. Forty-nine children whose I.Q.'s ranged from 135 to 190, with a median at 153 (Stanford Binet), gave about the same means on five of the Seashore phonograph tests as do normal children. And this occurred in spite of the fact that "their median intellectual ability was, therefore, at least that of the average adult."

In view of the norms established by Seashore for his rhythm test, this statement by DeGraff (8) is surprising: "It is usually believed that age influences rhythmic discrimination; that the older the person is, the better his rhythmic discrimination will be, but such is not the case, as a glance at the distribution table [his own] will show the distribution curve of the three age groups, namely, fifth-grade, eighth-grade, and adult, reveals an unusual degree of similarity. Although the adult has the advantage of maturity, enabling him to comprehend and concentrate on the test, his score on the rhythm test is only slightly higher than that of the fifth- and eighth-grade school child."

3. Training

Among the earliest experiments on training of pitch discrimination were those of H. S. Buffum (27). He "classified 28 eighth-grade pupils in a preliminary fifteen-minute individual test for each one and then proceeded to train them by requiring 40 minutes of the most specific and intensive practice under favorable conditions for 20 successive days, recording the average achievement for each day. The result was greatly surprising. Instead of showing progressive improvement with practice, each child remained as rated in the preliminary test, with only two exceptions, and the average ability for the class was the same at the end of the 20 days of this intensive practice as at the beginning."

F. O. Smith (30) has done a somewhat similar piece of work with 476 children, which corroborates Buffum. He concludes: "The sensitiveness of the ear to pitch difference cannot be improved appreciably by practice. There is no evidence of any improvement in sensitiveness to pitch as a result of practice."[1]

F. B. Ross (23) found that the discrimination of time as measured in the Seashore test (old form, not the Columbia record) displayed a marked improvement in eight subjects when the test was repeated seven times on successive days.

In some as yet unpublished studies on musical college students, F. A. Wright found that the giving of the Seashore battery of tests daily for a week increased the average scores slightly, but not to any significant extent. In a number of cases the highest score was made at the middle of the week.

It has been my experience that many subjects with no formal musical training make higher scores in distinguishing major, minor, and mistuned intervals than do subjects with considerable training. According to R. Seashore (29), this is also true with motor rhythm. Klauer (18) put subjects through two months of intensive marching, clapping, beating time, etc., and failed to find a significant increase in scores on sensory rhythm discrimination over the scores made by untrained control subjects. However, her results are complicated in that her control group was retested in a much shorter time than the training group and was obviously atypical. That is, its mean percentile was considerably lower than her training group's, so that no accurate comparisons can be drawn. R. Seashore (29) reports a correlation of .82, P.E. .02, between

[1] The interpretation of results in such mass experiments as those of Buffum and Smith is, in my judgment, decidedly difficult, if not often misleading. Certainly, an equally important method of studying the effects of practice is to confine one's effort to drilling a competent, though unmusical, adult under laboratory conditions which permit some measure of qualitative analysis of what takes place. If this is done, as I myself showed years ago in a study of pitch discrimination (*Amer. Jour. of Psych.*, 14: 1903, 561-568) not listed in the bibliography here, the effect of systematic training of an unmusical person is a curiously complex one—there appears marked improvement under certain conditions, along with none at all under others. Without going into all the details, suffice to say that I was able to train my subject to discriminate correctly pitch differences of less than three vibrations (a perfectly laudable achievement), but could at will alter conditions slightly and cause her to fail utterly to discriminate pitch differences, not of a few *vibrations*, but of *more than two octaves*. What is one to conclude from that? Can the unmusical be trained or can they not?—*Editor*.

scores on the sense of rhythm test and learning rhythm. The subjects in this experiment comprised those who stood in the highest and the lowest 10 per cent on the sensory rhythm test.

Seashore and Mount (28) report correlations between musical, instrumental or vocal training (as judged by questionnaires) and musical capacities (tested by the Seashore tests) which range from .39 down to .03. Correlations between musical environment (as judged by questionnaires) and capacities ranged from .30 to —.03.

In an interesting study by F. M. Brennan (4) training was found to correlate significantly (.40 or more) with pitch and rhythm discrimination, tonal memory, imagery, simple reaction, and intensity control. The training score was the actual number of half-hour lessons the student had had. It is well to bear in mind that causality is not necessarily expressed in such a correlation relationship. As Miss Brennan says, "In other words, his tonal memory need not be good because he had received a certain amount of training, but because he is possessed of a keen tonal memory, he naturally gravitates toward thinking and working with tones."

Using a similar training score criterion, De Graff (8) found correlations between it and rhythm discrimination of .09 with 464 adults, of .10 with 282 eighth-grade pupils, and of .21 with 272 fifth-grade pupils.

According to the work of Vance (35), pitch discrimination is keenest at 128, 256, and 512 d.v.—just the octaves most used in music. In fact, the musically trained subjects had better 'musical ears' at these pitches than the non-musical, but were poorer than they were at higher and lower pitches.

Under the heading of absolute pitch are listed such different phenomena as the ability to name a note played on one's own piano or violin, sung by a particular voice, played by almost any piano or violin or sung by another's voice, played with relatively pure tones, memory of a certain tonal frequency, etc. It is quite obvious from the many studies of this subject that absolute pitch can to some degree be acquired. In fact, it is taught by certain teachers, although other instructors consider it a doubtful aid to musical achievement and so do not encourage its development. It seems equally certain that some children display a marked degree of this

ability without having had any formal training. This fact has led certain writers to consider the ability as inherited—a belief which has little to substantiate it, but which cannot be proved faulty until more careful analyses of the early musical environments of these children are attempted (Gough, Ref. 12).

Prior to the work of P. F. Swindle (33) and M. F. Meyer (20), musical and psychological theorists held pretty generally that rhythm was an instinct. It is present to some degree in all races of men and thus meets the criterion of universality. If it is inherited, however, how does it happen that one finds two-point (two-step), three-point (waltz), four- and six-point rhythms in human activity, but no five- or seven-point rhythms? (There are some exceptions.) Swindle and Meyer both found it possible to train subjects so that they eventually 'felt' these unusual rhythms as intensely as the commoner four and six. They did this by forcing the subjects to beat out these rhythms day after day. Swindle concludes: "All experimental evidence seems to point to the conclusion that rhythm is acquired by each individual, and that it is not inherited. Biological conditions—for example, the anatomical fact that we are two-footed, two-handed, and generally two-sided, not three-cornered or starfish-like beings—are favorable for the development of those rhythms which have usually been considered to be instinctive, while the other rhythms can be acquired only under special, somewhat artificial conditions."

C. E. Seashore and others still hold that rhythm is an instinct, and that the innate capacity in this function can be measured. It seems to me that there is a way to reconcile partially these conflicting views. Might it not be that, although the specific rhythms one possesses must be acquired, one individual may profit more and faster from practice in rhythms than another? These individual differences may be due to nature or nurture—no one knows which—but they have musical significance. Rhythm tests would then be achievement tests.

Seashore (27) states that training undoubtedly favors improvement in the discrimination of consonance, but that the improvement is not so great as is commonly supposed. Strangely enough, he maintains that imagery is a special gift and that any increase in this ability is apparent, but not real. As the analysis of imagery

must be introspective, and so personal, I see no way of lifting this question from the realm of pure opinion. As to the effect of training on tonal memory, Seashore declares that "those with a large capital will improve, not only in proportion to their capital, but probably in increasing proportion for the larger capital."

Seashore summarized the effect of training on pitch as follows:

> At this preliminary stage, the following general conclusions may be stated:
>
> (1) When the proximate physiological threshold has been reached, practice is of no avail. (2) So long as a cognitive threshold prevails there is prospect of improvement by practice to the extent that the cognitive is above the physiological limit. (3) This improvement is usually very rapid, often immediate and can usually be traced to the acquisition of knowledge, through experience or information about what pitch is, as distinguished from other attributes of tone. (4) In the majority of cases it is possible for the ingenious experimenter to discover the proximate physiological threshold to a fair degree of certainty in a well-planned half-hour individual test, or in one heterogeneous and one homogeneous group test of one hour each; and, for most of the cases in which this fails, the same tests demonstrate positively that the threshold is only cognitive. (5) The possibility of reaching the physiological limit in a single test depends to but very slight extent upon whether or not the person tested has had musical education; it is mainly a matter of expert skill and ingenuity on the part of the experimenter. (6) Ordinarily musical education is not effective as a means of improving pitch discrimination.

The same author cites three lines of evidence to show that intensity discrimination is elemental:

> *Evidence: extended training.* Fourteen observers were selected regardless of initial ability and subjected to a prolonged series of specific and intensive practice of from fourteen to thirty periods of one half hour each, and a record was kept of performance from day to day. The general conclusion from this experiment was that systematic practice under the most favorable conditions did not result in a change of the record. On the whole, the achievement of these fourteen observers was practically the same at the end of the practice periods as at the beginning.
>
> *Evidence: the case of the blind.* It is generally supposed that, since the blind guide themselves by hearing so extensively, they must have more highly developed sensitiveness to

sound. To test this we took fifteen students in the School for the Blind who had been blind from early infancy and were normal in other respects, and had acquired skill in various kinds of work and ability to orientate themselves without sight. We compared the achievements of these with the achievements of fifteen seeing high-school pupils selected under similar conditions, and the result of this, the most careful test of its kind, was to show that the average for fifteen blind and fifteen seeing persons is approximately the same. This means that the intensified training which the blind have had since childhood has not tended on the whole to make them keener in this specific sensory capacity.

We found the same true of the hearing of direction, the sense of pitch, the sense of touch, and the sense of weight. This goes to show that the fundamental capacities of the senses are early developed to their maximum and that further development takes place not in the fundamental capacities but in the use of these capacities in more complex forms.

Evidence: the case of musical training. To these striking testimonials to the elemental character of intensity discrimination, we may add the records of children of the seventh and eighth grades who were classified on the basis of the amount of training, as above, and on the basis of intensity discrimination. In this we find a trace of agreement in the form of a tendency for those who have had extensive musical training to rank higher than those who have not. But the tendency is slight and is easily accounted for on the theory that those who have a keen appreciation of sounds have gravitated toward the opportunities of musical training.

A portion of the argument between the Seashore school and the musicians could be settled, I think, by distinguishing certain concepts carefully. Take the terms 'capacity' and 'ability.' As defined by Seashore (27), "The term 'capacity' has reference to the inborn or native power; the term 'ability' is used to designate acquired skill in the use of a capacity." It should be obvious, then, that the usual aim of the tester is to measure capacity, but he should hesitate to label his test scores measures of 'capacity,' as he rarely knows the exact extent to which environment has functioned. When he states that one's musical ear cannot be made keener, he is usually referring to the capacities in pitch, intensity, etc., and he may or may not be correct. The musician, however, well knows that, after training, a student can more easily tell when he is on

the key. This may be merely a shift from cognitive to physiological limit, or it may mean the entry of some other factor.

Again, take the term 'training.' (1) Training may mean merely the giving of a test a number of times, as in the studies of Wright, Ross (23) and others. (2) Training may mean repeated testing plus informing the subject which of his responses are correct and which incorrect. (3) Training may include explanations of the concepts involved, cues as to procedure, and the like. (4) Training may mean simply the accidental stimulation of the organism over long periods of time, with no formal instruction. Or (5) it may mean the formal training of an individual for a considerable period. Data on the effects of training will have different meanings as the meaning of training varies. Investigators should be specific in their descriptions of the training factor.

Certain of the psychoanalytic school hold that the musically talented are those who are expressing part of their libido in musical activities. Adler (1) points out that both Beethoven and Mozart had deformed ears. "I should only like to emphasize that to me the impression of the connection between a beautiful singing voice and organ inferiority which I have outlined is a definite one, and that I have considerable material to support it."

Conclusion

From the data considered in this paper, one is practically forced to conclude that the question of musical heredity is still a moot one. While genetics can illuminate the issue to some extent, I doubt that its contributions can ever be conclusive. With human stocks it is difficult to make very reliable Mendelian counts. And, even where musicality can be demonstrated as prevalent in certain family lines for generations, how can the factor of social inheritance be completely separated from that of biological inheritance?

Along the lines suggested by Seashore in his criterion of 'elemental' tests, the prospects for worthwhile work seem much brighter, and at least a beginning has been made.

Bibliography

1. Adler, Alfred. *Organ Inferiority and Its Psychical Compensation*, pp. 27-29.
2. Bartsch, K. *Zeit f. angew, Psych.*, 17: 1920, p. 144.
3. Bernfeld, S. *Archiv f. die ges. Psych.*, 34: 1915, p. 235.
4. Brennan, F. M. *Psych. Monographs*, 36: 1926.
5. Copp, Mrs. E. F. *J. of Heredity*, 1916, pp. 297-305.
6. Cox, C. *Genetic Studies of Genius*. Vol. II, 1926.
7. Davenport, C. B. *Heredity in Relation to Eugenics*. New York, 1911, p. 48.
8. De Graff, L. H. "The norms of sensitiveness to rhythm." *State University of Iowa Thesis* (M. A.) 1924.
9. Drinkwater, H. *J. of Genetics*, 5, pp. 229-241.
10. Galton, Sir Francis. *Hereditary Genius*. 1869.
11. Gaw, E. A. A monograph from Mills College. November, 1925.
12. Gough, Evalyn. *Archives of Psych.* 47: 1922, p. 94.
13. Hansen, C. F. *Psych. Monographs*. 31, 1: 1922.
14. Haecker, V. and Ziehen, Th. *Zur Vererbung und Entwicklung der musikalischen Begabung*. Leipzig, 1923.
15. Hollingworth, Leta. *J. of Ed. Psych.*, 17, 2, 1926, pp. 95-110.
16. Holmes, S. J. *The Trend of the Race*. Ch. III.
17. Hurst, C. C. *Eugenics Review*, 4: 1912, pp. 20-21.
18. Klauer, N. J. "The effect of training in rhythm upon rhythmic discrimination in the intermediate grades." *State University of Iowa Thesis* (M. A.), 1924.
19. Koch and Mjöen. *Zeit. f. Psych.*, 99: 1926.
20. Meyer, Max F. *Psychology of the Other-One*. Columbia, Mo., 1922.
21. Mjöen, J. A. *Eugenics Review*, 17: 1926, p. 242.
22. Révész, G. *The Psychology of a Musical Prodigy*. New York, 1925.
23. Ross, F. B. *Psych. Monographs*, 16: 1914.
24. Rupp, Hans. *Zeit. f. angew. Psych.*, 9: 1914.
25. Schüssler, H. *Zeit. f. angew. Psych.*, 11: 1916, p. 136.
26. Schüssler, H. *Zeit. f. angew. Psych.*, 19: 1921, p. 401.
27. Seashore, C. E. *The Psychology of Musical Talent*. 1919.
28. Seashore and Mount. *Psych. Monographs*, 25: 1918.
29. Seashore, Robert H. *Psych. Monographs*, 36: 1926.
30. Smith, F. O. *Psych. Monographs*, 16: 1914.
31. Stanton, H. M. *Psych. Monographs*, 31: 1922.

32. Stern, W. *Die Differentielle Psych.* 1911, p. 265.
33. Swindle, P. F. *Am. J. Psych.*, 24: 1913, p. 180.
34. Vaerting, M. *Zeit. f. Psychotherapie u. med. Psych.*, 7, pp. 120-7.
35. Vance, T. F. *Psych. Monographs*, 16: 1914.
36. Voss, G. *Dtsch. Zsch. f. Nervenhk.*, 83: 1925, pp. 249-263.
37. Wallace, Wm. *The Musical Faculty.* London, 1914.
38. Watson, J. B. *Behaviorism.* New York, p. 82.
39. Weaver, A. T. *J. of Appl. Psych.*, June, 1924.

CHAPTER XVI

A SUMMARY OF LITERATURE ON THE DETERMINERS OF THE INTELLIGENCE QUOTIENT AND THE EDUCATIONAL QUOTIENT

BARBARA STODDARD BURKS
Stanford University, Palo Alto, California

GUIDE TO CHAPTER XVI

Purpose and Nature of the Summary

In compiling the bibliography for this summary, and in classifying the summarized material, the general arrangement of the Yearbook has been followed as closely as possible. I have attempted to include at least brief mention of all material to date (June, 1927) which bears on the same themes as the original studies published in this Yearbook, *i.e.*, material which has offered crucial evidence in favor of or against either nature or nurture as determiners of mental capacity or of school achievement. Studies which seem to me to have made the most important contributions have been starred.

In addition to this clear-cut experimental material, much other material has been summarized because of the tendencies it reveals, the questions it provokes, the historical importance it holds, or the significance it contains as factual background for future research. To illustrate, nearly all the data summarized in the section on racial and nationality differences; and most of the data on intelligence in relation to differences in locality, parental status or home environment; on family resemblance; and on constancy of intelligence, fall into this type, as well as a number of other studies under practically every rubric of the entire summary. The number of such ancillary studies here assembled exceeds the number of the crucial type; even so, no pretence is made of covering the literature on such topics exhaustively. If this had been done, the Yearbook would have had room for nothing else. The notes upon some of the chief studies of this type (which are rather briefer than the notes upon studies of the first type) will at least give the reader an orientation in a variety of closely allied fields.

The summaries themselves are merely condensed, non-critical abstracts of the original sources. However, a number of them contain material which cannot be adequately interpreted out of its original context. Because I chance to have had occasion to read these contributions with care, I have taken the liberty sometimes of inserting comments which indicate the limitations or points of special significance not immediately obvious. These comments are always labeled and separated from the summaries proper; they

favor one side of the nature-nurture question fully as often as they favor the other.

References have been classified according to the field of evidence with which they are concerned. Within the separate classes the reader may also find it interesting to note other more or less distinct lines of cleavage which correspond roughly to the chronological order of the contributions.

A. Factual Material Leading Towards Hypotheses

1. Qualitative studies. These point out merely the existence of phenomena, often in individual cases, which require explanation. Large-scale tabulations of the incidence of genius and of feeble-mindedness in families or in groups of defined occupational, cultural, or racial status are included in this group.

2. Quantitative studies. Material of this type provides the groundwork for analytical studies that aim to separate and evaluate nature-nurture influences, though it contributes nothing crucial in itself, as for example, studies of family resemblance and constancy of the I.Q.

B. Evaluation of Nature-Nurture Effects

1. Presence-or-absence studies. These reveal evidences of hereditary or of environmental forces, but fail to answer the question "How much?" Studies of differential resemblance of siblings, fraternal twins, and identical twins fall in this class, as do studies of increments in intelligence-accomplishment correlations through intensive pushing of accomplishment.

2. Quantitative studies. These attempt to establish either

(a) The *proportional* contributions of nature or nurture to deviations in ability under defined conditions, or

(b) The quantitative effects (in terms of measurable ability units) of defined nature or nurture influences.

There are very few studies of either type (a) or (b) to be found.

I acknowledge gratefully the kindness of Dr. Rudolph Pintner and of Dr. Raymond R. Willoughby who put at my disposal unpublished bibliographies dealing with certain sections of this summary. These bibliographies were of substantial help; but for whatever shortcomings this summary may have I alone am responsible, since many phases of

the problem treated here were not included in the compilations of either Dr. Pintner or Dr. Willoughby. Dr. Pintner, in addition, permitted the writer to consult, and to incorporate in this summary wherever possible, material worked up by him upon thegeneral contributions of intelligence testing to the nature-nurture problem. This valuable material was of help in numerous ways in compiling the present chapter.

For the reader who wishes to familiarize himself with the views of current investigators in the field, the following references dealing with theoretical discussions or critical summaries rather than with experimental data, and hence not summarized in the body of this compilation, may be useful:

Burt, C. "The inheritance of mental characters." *Eug. Rev.* 4: 1912-1913, 168-200.

Elderton, E. M. "A summary of the present position with regard to the inheritance of intelligence." *Biometrika,* 14: 1923, 378-408.

Freeman, F. N. *Mental Tests.* Cambridge, 1926, 503 pp. (Ch. 17, "Interpretation of intelligence tests.")

Gates, A. I. "The inheritance of mental traits." *Psych. Bull.,* 18: 1921, 358-365.

Holmes, S. J. *The Trend of the Race.* New York, 1921, 396 pp. (Ch. 5, "The inheritance of mental ability.")

McComas, H. C. "The heredity of mental abilities." *Psych. Bull.,* 11: 1914, 379-383.

Peters, W. *Die Vererbung geistiger Eigenschaften und die Psychische Konstitution.* Jena, 1925, 368 pp.

Pintner, R. *Intelligence Testing.* New York, 1923, 406 pp. (Ch. 20, "The inheritance of intelligence.")

Poyer, G. *Les Problèmes Généraux de l'Hérédité Psychologique.* Paris, 1921, 302 pp.

Spearman, C. "The heredity of abilities." *Eug. Rev.* 6: 1914-1915, 219-237.

Starch, D. *Educational Psychology.* New York, 1919, 473 pp. (Ch. 6, "The inheritance of mental traits.")

Thorndike, E. L. and others, *The Measurement of Intelligence.* New York, 1927, 616 pp. (Ch. 16, "The measurement of original intellectual capacity and of acquired intellectual ability.")

The reader's attention is also called to the following volume, which contains a bibliography much more complete upon certain phases of the problem than the one presented in this summary:

Holmes, S. J. *A Bibliography of Eugenics,* Berkeley, 1924, 514 pp.

I. Intelligence

1. Family Resemblance

a. *Ordinary Family Relationships.* The general topic of family resemblance has been given first place in this bibliography because it involves the fundamental human phenomena which instigated the nature-nurture problem itself, and because studies in the province of family resemblance promise the only possible ground-work for future investigations upon the hereditary *mechanism* of mental traits. With the present status of our knowledge, however, it would be conceded by almost every one that few family-resemblance studies have contributed anything crucial to the general problem as yet, since resemblance conceivably might be due either to nature or to nurture or to both in various proportions.

When going over the results of studies summarized here, the reader should take special note of the fact that none of the correlations reported in the literature has been corrected for attenuation. If the unreliability of tests and ratings had been allowed for by such corrections, the correlations would, of course, give truer values for the actual amount of resemblance present.

1. Cobb, Margaret V., and Hollingworth, Leta S. "The regression of siblings of children who test at or above 135 I.Q. (Stanford Binet)." *Jour. Educ. Psych.*, 16: 1925, 1-7.

Thirty-seven gifted pupils at P. S. 165, Manhattan, had an average I.Q. of 155 by the Stanford Binet. Thirty-four siblings of these pupils had an average I.Q. of 129. Thirty-three first cousins had an average I.Q. of 127, but the authors think these may be selected. Twenty parents had a median Alpha score of 142.

2. Dexter, Emily S. "On family resemblance beyond the first degree of relation." *School and Soc.* 19: 1924, 501-502.

See section on family resemblance in achievement for summary.

3. Downey, June E. "Standardized tests and mental inheritance." *Jour. of Heredity*, 9: 1918, 311-14.

The following is given with respect to tests of parents and children: "Certainly it is significant that of groups of parents and children recently tested by my pupils the combination of parents, both of whom made a record of superior mentality by the Stanford adult scale, gave 80 percent of children making a superior or very superior record; while

when only one parent gave a superior record or both made only an average one, the percentage of superior children was only 33. Our groups were too small to permit our drawing any dogmatic conclusions."

Comment: The number of the cases is not reported.

4. Elderton, Ethel M. "On a measure of the resemblance of first cousins in man." *Eug. Lab. Memoirs,* 4: 1907, 1-53.

Three hundred pairs of cousins were rated for intelligence by volunteers. The ratings yielded a mean square contingency coefficient of .34.

5. *Elderton, Ethel M. "A summary of the present position with regard to the inheritance of intelligence." *Biometrika,* 14: 1922, 378-408.

Much of the early Galton laboratory work is reviewed. The following discussions, however, apparently appear for the first time.

(1) Kate Gordon's *second* series of orphan sibling data reworked. Miss Gordon's pairs contained a number of Jews, Mexicans, Italians, etc. Elderton thought that Gordon's r of .61 for 216 pairs may have exceeded the r of .51 (by Pearson's computation) on the first 91 pairs because of greater variability, but on using Gordon's original data, she found the variability no greater.

Gordon had used no child more than once in pairing to obtain her r of .61. Elderton now paired the siblings in all possible ways, made the table symmetrical, and got .47 ± .03.

Finding an r of —.31 ± .03 between age and I.Q., she corrected each child's I.Q. for the 1.59 drop per year, and computed r by the same method as above, getting .54 ± .02.

Partialling out C. A. in the ordinary way, by Gordon's method of pairing, older vs. younger, r was .58 ± .03.

Using all pairs, and C. A. constant, r was .54 ± .02.

Using all pairs, C. A. constant and table symmetrical, r was .53 ± .02.

Comment: The ages and length of time in orphanages are not mentioned here. The possibility of *early* home influences keeps the data from being entirely crucial, but the study is none the less very signficant.

(2) Dr. Drinkwater's data from a high class and a low class school. The numbers of siblings tested by Stanford Binet (and by Binet-Simon in the case of some of the younger ones) are not noted, but the P.E.'s of r's are reported. Results are given separately for the two schools, but just the combined results are summarized here. A negative r was found between intelligence and age in the case of girls but not in that of boys—an outcome that is attributed to selection. Ages and grades are not given. The undesirability of results based upon two kinds of tests is pointed out.

The r's for teachers' estimates are apparently computed by a broad category method (5 categories).

	I. Q.		Teachers' Estimate	
	r	P. E.	r	P. E.
Brother-brother	.49	.04	.49	.04
Sister-sister	.42	.04	.58	.03
Above, corrected for age	.38	.04	.54	.04
Brother-sister	.53	.02	.49	.02
Above, corrected for age	.53	.02	.50	.02

(3) Some data from Edinburgh are reported which show r's of − .07 to + .16 between the intelligence of children and such factors as drinking of the father, drinking of the mother, morality of the parents, physical condition of parents, economic condition of the home (.10 for boys and .16 for girls), overcrowding.

Comment: Drinkwater's data and the Edinburgh data appear useful only in establishing relationships. They tell nothing as to which is cause and which effect.

6. Goddard, H. H. *Feeble-Mindedness: Its Causes and Consequences.* New York, 1914, 599 pp.

From 164 family histories of children at the Vineland Training School for Feeble-Minded who were considered to be deficient through *hereditary causes,* field workers' reports upon the probable mental status of about 1700 individuals were analyzed. When members of the parent generations were classified according to probable Mendelian types of mating (upon the assumption that feeble-mindedness is a recessive *unit character*) the incidence of feeble-mindedness among the offspring fell approximately into Mendelian ratios. While the author is cautious in accepting a simple Mendelian explanation as established, he nevertheless states that the conclusion is justified that feeble-mindedness *behaves as if* such an explanation were valid.

Comment: There are significant deviations from the 'expected' Mendelian ratios in the case of several of the postulated types of 'Mendelian matings,' and among matings of two feeble-minded individuals, six out of 482 offspring are 'normal.' This latter could not occur, regardless of chance, if feeble-mindedness were actually a recessive unit character. But this discrepancy by itself does not entirely invalidate the unit-character hypothesis as applied to feeble-mindedness, since (a) some of the feeble-minded parents of the 482 offspring in question might be suffering from a secondary (*i.e.,* acquired and non-transmissible) form of deficiency; and (b) the classification of some individuals as 'normal' or 'deficient' may be in error. A more serious criticism of the conclusions is that, regardless of any approximation to Mendelian ratios which the types of offspring from given matings may offer, the hypothesis that deficiency, when it occurs below a certain arbitrary level upon a *continuous distribution* of human intelligence, is a unit character (*i.e.,* is

due to a single gene) is logically untenable. It can be shown (see discussion in Part I, Chapter II, of this Yearbook) that the ratios of feeble-minded and normal offspring found by Goddard could be explained about as well by a hypothesis that is fairly well in accord with the observed facts of human distribution as by the untenable hypothesis of a single recessive gene for feeble-mindedness.

7. *Gordon, Kate. "The influence of heredity in mental ability." *Children's Dept. Calif. State Board of Control, 4th Bienn. Report.* 1918-1920.

See extension of this report summarized under Elderton, No. 5.

8. Gordon, Kate. "Report on psychological tests of orphan children." *Jour. of Delinquency,* 4: 1919, 46-56.

See extension of this report summarized under Pearson, No. 13.

9. Hart, Hornell. "Correlations between intelligence quotients of siblings." *School and Soc.,* 20: 1924, 382.

In connection with test work of the Iowa Child Welfare Research Station, Army Alpha, National, or Stanford Binet tests were obtained from the following groups of siblings:

A. Children aged 10 to 15 in Davenport, Iowa.
B. Children of farmers in Sumner Township.
C. Children attending the University schools, Iowa City.

When I.Q.'s were derived (occasionally from averages of combinations of tests), correlations between siblings were:

Group	r	P. E.	N
A	.45	.03	252
B	.46	.07	147
C	.40	.06	219

Comment: Averaging I.Q.'s from tests of varying types seems dubious. The ranges of the groups are not given.

10. Heymans, G., and Wiersma, E. Beiträge zur speziellen Psychologie auf Grund einer Massenuntersuchung. *Zeits. für* Psych., 42: 1906, 1-127, 258-301.

See account and reduction of data by Schuster and Elderton.

11. *Hildreth, Gertrude H. *The Resemblance of Siblings in Intelligence and Achievement,* N. Y., 1925, 65 pp. (Teachers College, Columbia University, Contributions to Education, No. 186.)

The author analyzed results of tests given by schools and by an orphanage. The groups tested were:

A. Four hundred fifty sibling pairs from 300 families in a city in Oklahoma apparently representative of average American cities. These children were not unselected, however, since the schools tried to test only bright and dull children for special-class work, and hence dealt mainly with dull children.

B. Three hundred twenty-five sibling pairs from 241 families. This group, who attended Horace Mann school, had a disproportionate number of bright children.

C. Two hundred fifty-three pairs representing 146 families. The records were obtained from a large Hebrew orphan asylum in New York City. Children in the asylum were received at any age above 5 years.

The range of I.Q.'s and ages for the three groups was as follows:

Group	I. Q.			C. A.		
	Q_1	Med.	Q_3	Q_1	Med.	Q_3
A.	72	88	103	9–11	11–10	13–3
B.	108	119	129	5–2	7–5	9–8
C.	85	96	105	8–0	9–11	11–7

Stanford Binet records were available for all (excepting a few Oklahoma cases which had been tested upon other Binet revisions). No tests were made especially for the investigation. For some of the Oklahoma and Horace Mann cases Stanford Achievement Test data were available. Care was taken to eliminate children thought to be sibs who were not actually sibs.

In the correlations all possible pairings were used, and the pairings were made at random—not oldest with youngest or brightest with dullest.

Correlations Between I. Q.'s

	Raw r		Age partialled out		
	r	P. E.	r	P. E.	N
Oklahoma	.63	.02	.47	.02	450
Horace Mann	.27	.03	.08	.04	325
Hebrew Orph.	.32	.04	.13	.04	253
Composite	.68	.01	.41	.02	1028

	Correlations Between E. Q.'s		Correlations Between A. Q.'s		N
	r	P. E.	r	P. E.	
Oklahoma	.58	.04	.27	.06	105
Horace Mann	.41	.06	.16	.07	83
Composite	.58	.03	.32	.04	188

To investigate the influence of environment upon the variability in I.Q., data from the orphanage were treated as follows:

	Mean I.Q.	S. D.	N
1. Unrelated children reared together:			
10– 25 percent of their lives..........	96	11.7	94
50–100 percent of their lives..........	93	13.5	94
2. Unrelated children reared apart.........	96	13.6	200
3. True sibs reared apart.................	93	11.4	156

True sibs reared apart yielded the correlations that follow:

	r	P. E.	N
All pairs..............................	.23	.07	78
One pair, each family....................	.24	.10	42
All pairs corrected for curtailed range......	.495	.06	78

Under just what conditions tests of the siblings reared apart were obtained is not reported. The average time apart is said to be four years. The separations occurred from birth to age 12.

The author concludes: "True siblings reared apart for part of their lives still resemble each other about as much as true sibling pairs reared together. Unrelated children reared together for part of their lives resemble each other in intelligence no more closely than unrelated children chosen from the population at random."

Comment: The data suggest an explanation based partly on heredity, but the author admits that the data are inconclusive. There is no way of evaluating the influence of pre-institutional life upon the sibling resemblances. On the other hand, the low correlations between siblings reared in an orphanage suggest some environmental effect. It would be of interest to know whether the correlation of .32 for orphanage siblings (or of .13 when C. A. is partialled out) would rise if a correction for range were applied.

12. Madsen, I. N. "Some results with the Stanford Revision of the Binet-Simon Tests." *School and Soc.*, 19:1924, 559-562.

Sixty-three pairs of siblings tested by students of the author yielded a correlation of $.63 \pm .05$. The author writes: "Undoubtedly the correlation would have been larger if all errors in giving and scoring the tests could have been eliminated."

Comment: The range of the subjects was not given. Unless unusual selective influences were operative, .63 is a higher r than genetic resemblance of siblings should yield. This high value might possibly be due, however, to the inclusion of children older than the limit below which the Stanford Binet gives zero correlation with age, in which case,

if the C.A.'s of siblings were correlated, the mental-test correlation would be spuriously raised.

13. Pearson, Karl. "Inheritance of psychical characters." *Biometrika*, 12 : 1918-19, 367-372.

The article opens with the following comments: "In view of the papers that have been published on the inheritance of intelligence, it is strange that there should still remain any doubt that psychical characters are inherited at the same rate as physical characters. But having regard to the existence of that doubt, any material bearing on the point deserves special recognition and emphasis."

Taking Gordon's first 91 orphanage sibling pairs (who have a mean I.Q. of 93 and S.D. of 16.7), he renders the table symmetrical and gets an r of .51 ± .05, as compared with .53 obtained by Gordon. Apparently, her pairing was used, since he used her original table for his computation.

Comment: These data are of course more crucial than ordinary data on family resemblance, but resemblance due to home influence in early life of the siblings remains a speculative possibility.

14. Pearson, Karl. "On the laws of inheritance in man. II. On the inheritance of mental and moral characters in man." *Biometrika*, 3 : 1904, 131-190.

The correlations quoted [see next page] are based on ratings or measurements by members of the family or teachers. The numbers are so large that P.E.'s are negligible.

15. Pearson, Karl. "On the inheritance of the mental characters in man." *Proc. Royal Society London*, 69 : 1901, 153-155.

Comment: The following conclusion, as enunciated in this paper, is quoted because it illustrates the extreme, and seemingly unwarranted, confidence that is sometimes drawn from data on family resemblance. "Dealing with the means for physical and mental characters, we are forced to the perfectly definite conclusion that the mental characters in man are inherited in precisely the same way as the physical." See notes in this section on Spearman's criticism of Pearson. See also the foregoing summary of Pearson's 1904 article, setting forth in amplified terms the results of this brief preliminary paper.

16. Pintner, R. "Training students in group intelligence testing." *Jour. Educ. Research*, 9 : 1924, 271-280.

Three groups of siblings, aggregating 106 pairs, yielded correlations between .24 and .42.

Comment: The type of tests employed was not stated.

[To be read with Item 14 on p. 258.]

ADULTS FROM 1,000 FAMILIES

Character	Correlation		
	Brothers	Sisters	Brothers and Sisters
Stature	.51	.54	.55
Span	.55	.56	.53
Cubit	.49	.51	.44
Eye color	.52	.45	.46
Mean	.52	.51	.49

SCHOOL OBSERVATION SERIES

Physical Characters	Correlation		
	Brothers	Sisters	Brothers and Sisters
Health	.52	.51	.57
Eye color	.54	.52	.53
Hair color	.62	.57	.55
Cephalic index	.49	.54	.43
Head length	.50	.43	.46
Head breadth	.59	.62	.54
Head height	.55	.52	.49
Mean	.54	.53	.51
Athletic power	.72	.75	.49

SCHOOL OBSERVATION SERIES—*Continued*

Mental Characters	Correlation		
	Brothers	Sisters	Brothers and Sisters
Vivacity	.47	.43	.49
Assertiveness	.53	.44	.52
Introspection	.59	.47	.63
Popularity	.50	.57	.49
Conscientiousness	.59	.64	.63
Temper	.51	.49	.51
Ability	.46	.47	.44
Handwriting	.53	.56	.48
Mean	.52	.51	.52

17. Pintner, R. "Mental indices of siblings." *Psych. Rev.*, 25: 1918, 252-255.

Yule's coefficient-of-association technique, applied to mental indices of over 180 pairs of siblings, gave a value of .39. Indices were based upon results from 6 tests ranging from opposites to cancellation. The method of deriving the indices is not stated.

18. Rensch, Grace E. *A Study of 365 Pairs of Siblings with Reference to the Influence of Heredity on Mental Ability.* (Stanford M. A. Thesis, 1921, 35 pp.)

The Stanford Binet or the Terman Group Test was applied to groups of Palo Alto and San Jose siblings ranging from the first to the twelfth grade. Barr occupational ratings were also obtained for the fathers or for the mothers if the fathers' occupations were unknown. The average of correlations between siblings of various sub-groups was .45. The average of correlations between I.Q.'s of children and Barr ratings of fathers (or mothers) was .33.

Comment: Nothing was said of the possibility that occupational 'stratification' of various races represented by the subjects might partly account for the rather high correlations between Barr ratings and I.Q.'s. The justification for combining in single correlations I.Q.'s based upon two different tests is doubtful.

19. Schuster, E., and Elderton, Ethel M. "The inheritance of psychical characters." (Summary of Heymans and Wiersma.) *Biometrika,* 5:1906-1907, 460-469.

A summary and reworking of data of Heymans and Wiersma is presented. Three thousand Dutch physicians were sent a questionnaire asking for family ratings upon 90 psychial characters. More than 400 of these were returned. Correlations between fathers and sons and daughters, and between mothers and sons and daughters averaged between .20 and .30.

Comment: The study has historical importance, but the lack of standardization in ratings by the judges renders its correlations ambiguous.

20. Spearman, C. "The proof and measurement of association between two things." *Amer. Jour. Psych.,* 15:1904, 72-101.

It is pointed out that correlations between psychic characters based upon teachers' judgments, as reported in Pearson's work, must be attenuated by unreliability of judgments. The reliability of such judgments is not usually above .64, which would bring Pearson's reported correlations of .52 up to .81. However, attentuation may be partly balanced by 'dilation' owing to restrictions of cases to brothers attending the same schools. Those with different qualities are perhaps at sea, working in offices, etc.

21. Woods, F. A. *Mental and Moral Heredity in Royalty.* New York, 1906. 312 pp.

Members of the royal families of Europe, extending from the present back to about the 16th century, and even to the 11th century (Spain

and Portugal), were studied. The discussion was chiefly confined to individuals who had reached thirty or over.

By consulting biographical dictionaries and lexicons, Woods graded all individuals upon intellect and moral quality on a scale descending from 10 to 1. The ratings fell in a fairly normal curve.

Mental and moral ratings on 608 individuals correlated $.34 \pm .04$ by Pearson's four-fold method (division at the mean).

The following four-fold correlations were found, and compared with what Woods erroneously takes to be the theoretical r's.

	Mental Ability		
	N	Obtained r	Theoretical r
Father and offspring	504	.31	.30
Grandfathers and offspring	952	.16	.15
Great grandfathers and offspring (apparently refers to the maternal grandfather plus two great grandfathers)	1,179	.15	.075

It is pointed out as a significant fact that the correlation is as high on the maternal side as on the paternal, although maternal grandfathers must represent different courts, and hence different *milieus* from those in which the offspring are reared.

The conclusion is that, since the grandparents . . . "furnish their full quota of resemblance," although the environment is likely to be different, "we must then consider the resemblances to be brought about through the germ-cells alone."

It is shown that inheritors of thrones whose position might help them to gain intellectual distinction appear in the top levels no more frequently than younger brothers.

Comment: Environment is not ruled out in any rigorous sense, although the environment of the entire group is no doubt superior and somewhat uniform. Thus, the results almost surely demonstrate some influence of heredity, since differences in environment could not account for the large differences in ability.

b. *Twins.* The fact that twins resemble one another more than ordinary siblings do, and that identical twins resemble each other more than fraternal twins do, is strong presumptive evidence in favor of heredity. However, we have no way of telling whether or not part of the greater resemblance in twins may be due to a greater similarity in the environment provided for twins than for ordinary siblings. Similarly, in the comparison of identical and fraternal twins, we do not know whether the environment of the

former may not be more similar than that of the latter. Furthermore, even if the environment for siblings, and for both types of twins could be assumed equivalent, it would be difficult to say just what interpretation to put upon the difference in correlation between identical and fraternal twins; *e.g.*, how much has heredity or environment affected scores if the correlation for fraternal twins is .50 and that for identical twins is .85?

22. Galton, Francis. *Inquiries into Human Faculty.* London, 1883, 387 pp. (History of twins, pp. 216-243).

Data were collected by the author through written inquiries and reports from 35 twin pairs of close similarity and 20 pairs of early dissimilarity.

The study is not reported statistically, but many excerpts from communications are reproduced. In the "similar" group a dozen cases are mentioned in which some sickness apparently caused a pair to grow dissimilar in some traits (intelligence?). In the few cases in which twins similar when young grew diverse "where no external cause can be assigned . . . we may feel sure that it must be chiefly or altogether due to a want of thorough similarity in their nature. . . . In some cases the resemblance of body and mind had continued unaltered up to old age, not withstanding very different conditions of life."

In the "dissimilar" group, the author has not a "single case in which . . . correspondents speak of originally dissimilar characters having become assimilated through identity of nurture."

23. *Merriman, Curtis. *The Intellectual Resemblance of Twins.* (Psych. Monographs, 33: 1924, No. 5, pp. 58).

The entire twin population of Grades I to VIII in certain cities was studied. Data were secured for 105 pairs on the Stanford Binet, 76 pairs on Army Beta, and 143 pairs on the National Intelligence Test. A summary of the chief correlations on these three tests is presented herewith.

	Binet			Beta			N. I. T.		
	r	P.E.	N	r	P.E.	N	r	P.E.	N
All	.78	.02	105	.84	.02	76	.89	.01	143
All 5–9	.81	.03	47	.78	.05	28	.80	.03	54
All 10–16	.76	.04	58	.66	.05	48	.88	.02	89
Like sex	.87	.02	67	.91	.02	45	.92	.01	92
5– 9	.88	.03	29	.92	.02	16	.95	.01	31
10–16	.86	.03	38	.84	.04	29	.86	.02	61
Unlike sex	.50	.08	38	.73	.06	31	.87	.02	51
5– 9	.77	.06	18	.52	.15	12	.75	.07	23
10–16	.30	.14	20	.64	.09	19	.83	.04	28

The author points out that there are higher correlations between like-sex pairs than between unlike-sex pairs, that there are no significant differences between the correlations of twins of different ages, that special study of pairs reported as "similar" shows even greater likeness than that existing in the total group of like-sex pairs, and that correlations between twins of unlike-sex are about equivalent to those usually found for ordinary siblings. An explanation of results in terms of heredity is suggested.

Comment: Apparently the scores with the Beta and the National tests were not corrected for age, which, of course, would spuriously raise these correlations. (See summary of critique of this study by Eugene Shen, No. 26.)

24. Muller, H. J. "Mental traits and heredity." *Jour. Heredity,* 16: 1925, 433-448.

Two young women of 30, identical twins, who were separated at two weeks of age and did not see each other until age 18, were tested by the Army Alpha, the Otis Self-Administering Test, and several personality tests. Physical measurements were also made of the twins. Their scores on Alpha were 156 and 153; and on the Otis test, 64 and 62. Though both had been brought up in the country, one had had only four years of formal schooling, including 9 months in business college; while the other had gone through the high school and had taught school until her marriage 9 years ago.

Their similarity on the mental tests was most striking, on the separate test elements hardly less than upon the total scores. On the other hand, their performances on the personality tests showed no unusual similarity.

Comment: The results with the mental tests have decided significance; but those on the personality tests are hard to interpret because of the unreliability of the tests used.

25. Popenoe, Paul. "Twins reared apart." *Jour. Heredity,* 13: 1922, 142-44.

These are the twins discussed in detail by Muller in 1926. This account contains no test data, but presents brief histories of each twin and anecdotal reports concerning simultaneous activities of each without the knowledge of the other—as sickness, bobbing their hair, etc. Friends cannot tell their voices apart—or their faces, usually.

26. Shen, Eugene. "The intellectual resemblance of twins.' *School and Soc.,* 21: 1925, 601-602.

Merriman's conclusion that environment appears to make no significant difference in the amount of twin resemblance is discussed critically. Merriman's data were not entirely crucial because:

(1) Reliabilities were not reported so that differences between pairs could be separated into true and chance differences (by correction for attenuation).

(2) Age was not allowed for in the group tests—thus giving rise to a spurious correlation.

(3) The range of the scores for different groups was not reported; consequently, the r's are not strictly comparable. Shen has made the data for the various groups comparable by computing the I.Q.'s and standard errors of estimate in each case from data in Merriman's appendix. The resulting standard errors of estimate show a quite consistent increase of 3 points from younger to older groups in all twin classes.

This might be explained (but is not through Merriman's treatment) by (1) decreasing reliability of the I.Q., (2) increasing variability of the I.Q., or (3) increasing effect of environment from lower to upper ages. If (1) and (2) are not sufficient to account for it, then (3) must.

27. Thorndike, E. L. *Measurements of Twins.* (Columbia Univ. Contrib. to Phil. and Psych., 13: 1905, 64 pp.)

A battery of six mental tests (such as cancellation and opposites) and of eight physical measurements was applied to 50 pairs of twins aged from 9 to 15. The average correlation, corrected for attenuation, of scores on the six mental tests (in terms of age norms) was .78. The average of the correlations for the eight physical traits (measured in terms of age norms) was .76. The author finds that the young twins of the group are as much alike as the older ones, that the twins resemble each other more than the ordinary siblings who took the test, and that resemblance is as strong in relatively untrained capacities as in trained ones. He thereby concludes that the similarities are due to inborn capacity.

Comment: Data are also presented which the author believes demonstrate that all twins are of only one type (rather than of both identical and fraternal types), but the arguments rest upon such an uncertain statistical basis that they can not be accepted in the light of more recent studies.

c. *Incidence of Genius or Feeble-mindedness in Family Lines.* This section is of historical importance chiefly. The studies summarized here call attention in a forcible way to family resemblance, but fail to furnish quantitative statements of such resemblance.

28. *Brimhall, D. R. "Family resemblances among American men of science." *Amer. Nat.*, 56: 1922, 504-547, and 57: 1923, 137-152, 326-344.

Questionnaires obtained from 956 of Cattell's starred group of 1000 men of science, combined with information from biographical sources, yielded the following data:

Brothers of these men are 72.8 times as likely to be in *Who's Who* as are brothers of the generality, and sisters are 70.8 times as likely to be there as are women of the generality. A brother is twice as likely to be distinguished as a father, four times as likely as an uncle, and six times as likely as nieces and nephews.

If relatives are taken no more remote than cousins and if both the father and the mother are omitted, there are 117 distinguished relatives on the paternal side, of whom 30.8 percent were distinguished in science; and 110 on the maternal side, of whom 36.3 percent were distinguished in science.

Comment: This latter result is perhaps the most significant one of this study, since the maternal influence was equal to the paternal influence, even though the economic status of the immediate families must have been chiefly determined by the fathers.

29. Cattell, J. McKeen. *A Statistical Study of American Men of Science: a Biographical Directory.* New York, 1921, 27 pp. 3d Edition. Reprinted from *Science*, N. S., 24: 1906, 658-665, 699-707, 732-742.

American men of science had far more professional men among their fathers than an equivalent number of men of the generality. (See Brimhall's study upon the heredity of men from Cattell's group.)

30. Danielson, Florence H., and Davenport, C. B. *The Hill Folk.* L. I., N. Y., 1912, 56 pp. (Eugenics Record Office Memoir, No. 1.)

A family connection called the Hill Folk abounded in feeble-minded and alcoholic members.

31. Dugdale, R. L. *The Jukes; a Study in Crime, Pauperism, Disease, and Heredity.* New York, 1877, 120 pp.

Of 1200 descendants of a shiftless truant born in 1720, only 20 had learned a trade. The genealogy was replete with paupers, murderers, and enemies of society.

32. Estabrook, A. H. *The Jukes in 1915.* Washington, 1916, 85 pp. (Carnegie Institution of Washington, Publication No. 240.)

The author brings the study of the Jukes made by Dugdale in 1877 up to date. The family stock had not improved.

33. Estabrook, A. H., and Davenport, C. B. *The Nam Family.* L. I., N. Y., 1912, 85 pp. (Eugenics Record Office Memoir, No. 2.)

In the family Nam, 90 percent of the men and 80 percent of the women were alcoholic or shiftless.

34. Galton, F. *Hereditary Genius.* London, 1869, 390 pp.

Nine hundred and seventy-seven eminent men each defined as the most eminent of 4000 random persons had 535 eminent relatives as close as grandfathers or grandsons. Nine hundred seventy-seven men selected by chance would have only four such relatives.

35. Galton, F. "Distribution of successes and of natural ability among the kinsfolk of fellows of the Royal Society." *Nature,* 70: 1904, 354-356.

Family schedules obtained from 200 Fellows of the Royal Society showed a large number of relatives as distinguished as the Fellows themselves. There was a consistent dropping off in numbers of such relatives as relationships became more remote.

36. Goddard, H. H. *The Kallikak Family.* New York, 1912, 121 pp.

Two family lines established by a soldier in the Revolutionary War were compared. One line established by a feeble-minded woman contained 480 direct descendants among which only 46 normal individuals were found. Among 496 direct descendants of the line established by a normal woman, all were normal with the exception of five.

37. Jörger, J. *Psychiatrische Familiengeschichten.* Berlin, 1919, 494 pp.

Gives account of the notorious family Zero.

38. Winship, A. E. *Jukes-Edwards: A Study in Education and Heredity.* Harrisburg, 1900. 88 pp.

Fourteen hundred descendants of the Jonathan Edwards family contained a preponderant number of college graduates, professors, and literary men. No mental defectives were traced in the family line. This line is compared with the notorious Jukes family.

39. Woods, F. A. "Heredity and the Hall of Fame." *Pop. Sci. Monthly,* 82: 1913, 445-452.

Biographical dictionaries list about 3500 eminent Americans. The average man has only one chance in 500 to have an eminent man for a relative as close as uncle or grandson. But one in five of the 3500 Americans has a relative among the 3500. Of the 46 Americans in the Hall of Fame, more than one in two has an eminent relative. The author concludes that something like one percent of the population is as likely to produce a man of genius as all the rest of the population.

d. *Animal Studies.* Since the environment for laboratory animals is as constant as it is possible to have it, family resemblance in animals almost surely indicates hereditary transmission of traits. Studies of learning capacity only are summarized below, since the few other studies in the literature are concerned with inheritance of temperament or of acquired characteristics, two topics not touched upon in the body of this summary.

40. *Bagg, H. J. *Individual Differences and Family Resemblances in Animal Behavior.* New York, 1920, 58 pp. (Archives of Psych., No. 43.)

One hundred eighty-three white mice and yellow mice were given one trial a day in a simple maze requiring only two choices. Seventeen trials were given, of which only the last 15 were used in the results.

The yellow group spent an average time of 83 ± 7 seconds, with an average of two errors. The white group spent an average time of 27.5 ± 2 seconds, with an average of 0.9 error a trial. "The yellow mice were thus found to take on the average at least three times as much time and to make twice as many errors as did the white mice.

Data upon maze records for six or seven generations lead the author to the conclusion that "there appears to be no particular resemblance among individuals of the same litter, as judged by their time-records for various tests." The results with mice of the white and yellow strains, however, suggest "a considerable difference among strains."

Comment: The study is of importance in demonstrating gross differences in maze-running ability for different strains of rats. The maze was too short, and hence too unreliable, and the character of inbreeding too uncertain to make an interpretation of the results by biometric or genetic methods possible.

41. Burlingame, Mildred. "Literature on the heredity of behavior traits in animals." *Psych. Bull.*, 24: 1927, 62-68.

Summarizes the chief studies briefly, including temperamental traits as well as of learning capacity.

42. *Tolman, E. C. "The inheritance of maze-learning ability in rats." *Jour. Comp. Psych.*, 4: 1924, 1-18.

By selective breeding two independent strains of rats were established, one 'bright' and the other 'dull.' From a parent generation of 41 males and 41 females, 9 pairs of bright rats and 9 pairs of dull rats were selected upon a rough pooling of time, errors, and number of perfect runs (on a maze 44 by 44 inches on the outside.)

Sixteen pairs from 18 litters were the parents of a third generation.

The animals were given 10 trials on successive days. Omitting runs 1 and 2, the reliabilities based upon odd versus even runs were about .38 for errors and .59 for time.

Correlations between mid-parent and mid-offspring for the three generations were as follows:

	Errors	Time	Perfect runs
P vs. F_1 (18)...................	.46	.30	.86
F_1 vs. F_2 (16)................	.32	−.27	.78

Means and variabilities of the 'bright' and 'dull' of the three generations were also reported and showed distinct differences between the bright and dull strains in the second (F_1) generation, but not in the third (F_2). The author suggests that the convergence in F_2 may be due "to the strict inbreeding which was practiced or . . . the respective ages at which the animals were tested."

Comment: The study shows distinct evidence for heredity, but leaves us in the dark as to the strength of influence of heredity. The correlations are ambiguous because they are based upon populations from the extremes of the normal distribution; and they are computed with the *averaged* scores of offspring as one variable, which spuriously raises them.

2. Intelligence and Social Environment

The bearing of studies in this field is indirect unless measured *changes* in environment are shown to accompany changes in mental status. Hierarchies of intelligence in social environments of various grades may or may not mean environmental influences; they seem at least as likely to indicate the influence of intelligence (of individuals or of their parents) upon environment.

a. *Occupational Status.*

43. Alexander, H. B. "A comparison of the ranks of American states in Army Alpha and in social economic status." *School and Soc.*, 16: 1922, 388-392.

Army Alpha rank orders of states (based on white drafted men) for 44 states are correlated with rank orders in other variables. Rank correlations are transmuted to r by table. The results follow:

The author concedes that the results may be interpreted either in favor of heredity or environment, but concludes that "Army Alpha appears as a test of what *has* been learned rather than what *can* be learned."

Alpha Rank Order With:	r	P. E.
Ratio of men in 1912 *Who's Who* to population of states in 1860	.79	.05
Percent of foreign-born whites in 1910 census (author finds this result surprising)	.65	.06
Percent urban population in 1910	.62	.07
Percent owning homes in states having over 50 percent rural population	.68	.07
Percent of farms owned by white men	.70	.05
Average wage for farm labor	.83	.03
Literacy (of whites)	.64	.06
Ayres school index in 1900	.72	.05

Comment: The fact that these high correlations were based upon group *averages* must, of course, be kept in mind. This fact causes the actual relationships to be greatly exaggerated. (See Chapter II of this Yearbook, on "Statistical Hazards.") The causal interpretation is unwarranted.

44. Barr, Frank E. *A Scale for Measuring Mental Ability in Vocations and Some of its Applications.* (M. A. Thesis, Stanford University, 1918, 90 pp.)

The author describes the derivation of a scale based on estimates of twenty judges upon the order of 100 occupations in the demands made on mental ability. The r of 10 against 10 judgments on the 100 occupations is .90 and .95 by two arrangements. P. E. is .01.

I.Q.'s of children in various schools (Palo Alto, Mayfield, Oakland, San Jose, Santa Clara) are correlated with Barr ratings of fathers on a 14-point division of the scale. On the Stanford Binet, r's at different schools are .69 ± .04; .41 ± .07. With lumped results of Binet and some kind of "reliable group test," Barr ratings correlate .50 ± .03 for 234 cases.

Comment: It seems possible that these r's may be raised by foreign elements in the population. It is a dubious procedure to lump I.Q.'s from two types of test.

45. Book, W. F. *The Intelligence of High-School Seniors.* New York, 1922, 371 pp.

The percentages of high-school seniors scoring above the group medians were tabled according to the occupational group of their parents with the following results:

Professional	60	Executive	54
Clerical	60	Day Laborer	47
Salesmen	56	Farmer	43
Artisan	55		

City children were also found to score higher than country children.

46. Clarke, E. L. *American Men of Letters.* New York, 1916, 169 pp. (Columbia University, Studies in History, Economics, and Public Law, No. 168.)

Of 666 American men of letters, 49 percent were descended from professional classes and only seven percent from mechanical, clerical, and unskilled labor groups.

47. Cox, Catharine M. *Genetic Studies of Genius, Vol. 2. The Early Mental Traits of Three Hundred Geniuses.* Stanford University, 1926, 842 pp.

The occupations of fathers and maternal grandfathers of 282 of the most eminent historical personages living between 1450 and 1850 were distributed on the Taussig Scale as indicated below:

Taussig Rating and Classification	Fathers		Maternal Grandfathers		
	Freq.	Percent	Freq.	Percent of 282	Percent of 182 reported
1. Professional and nobility	148	52.5	77	27.3	41.8
2. Semi-professional, higher business, and gentry	81	28.7	65	23.0	35.3
3. Skilled workmen and lower business	37	13.1	35	12.4	19.1
4. Semi-skilled	11	3.9	3	1.1	1.6
5. Unskilled	3	1.1	4	1.4	2.2
No record	2	0.7	98	34.7	....
Total	282	100.0	282	99.9	100.0

(See also the section on constancy of I.Q. for summary of other significant details of this study.)

48. Dashiell, J. F., and Glenn, W. D. "A reëxamination of a socially composite group with Binet and performance tests." *Jour. Educ. Psych.*, 16: 1925, 335-340.

About a dozen children in each of three groups were tested by the Stanford Binet and by six tests from the Pintner-Paterson Performance Scale with results noted herewith.

Groups	Binet I. Q.		Performance (points)	
	Av.	A. D.	Av.	A. D.
Children of faculty members	112	9.0	156	38.2
Town	100	7.1	103	27.1
Country	94	8.1	111	22.3

The authors suggest that the shift in rank of town and country groups on the performance scale may be due to "a difference in the nature of the two kinds of tests as respects the exact traits measured" or to the possibility that capacities of the town and country children are differentiated, "not by fundamental nature," but by "previous experience with the types of things called for in the examination." "As for the faculty group, it is clear that here is a distinctly abler type of child." The authors admit that the data are too few to be conclusive.

Comment: There is no explanation why the faculty group is equally high on both tests, while the other groups shift places.

49. Dexter, Emily S. "The relation between occupation of parent and intelligence of children." *School and Soc.*, 17:1923, 612-614.

Tests of 2782 children in Madison (Dearborn Test in Grades I to III and National Test in Grades IV to VIII) showed average I.Q.'s ranging from 115 for those from the professional classes down to 89 for those from unskilled laboring class.

Taking the average I.Q.'s of children of fathers of definite occupations, and the average Army Alpha scores of men of the same occupations from three camps, the rank order correlations were .74, .79, and .79.

Comment: These high correlations must, of course, be interpreted as based upon data from *group averages.*

50. Ellis, Havelock. *A Study of British Genius.* London, 1904, 300 pp.

One hundred three British men and women of genius were studied. They were descended predominantly from upper-class and professional families.

51. English, H. B. *An Experimental Study of Mental Capacities of School Children Correlated with Social Status.* New Haven, 1917. (Yale Psychological Studies, Psych. Monographs.)

Sixty-five percent of a group of lower-class children fell below the median of the total group (which included the lower-class children).

52. Haggerty, M. E., and Nash, H. B. "Mental capacity of children and paternal occupation." *Jour. Educ. Psych.*, 15:1924, 559-572.

In a New York mental-test survey, high-school and elementary-school pupils tested on Haggerty Delta 2 showed a consistent drop in median level from professional to unskilled laboring classes.

53. Pressey, S. L., and Ralston, R. "The relation of the general intelligence of school children to the occupation of their fathers." *Jour. Appl. Psych.*, 3:1919, 366-373.

When children were grouped according to the occupations of their parents, the following percentages scored above the median of the group: professional, 85; executive, 68; artisan, 41; laborer, 39.

54. Rensch, Grace E. *A Study of 365 Pairs of Siblings with Reference to the Influences of Heredity on Mental Ability.* (Stanford M. A. Thesis, 1921, 35 pp.)

See summary in section on family resemblance.

55. Sandiford, P. "Paternal occupations and intelligence of offspring." *School and Soc.*, 23: 1926, 117-119.

Tests of over 5000 high-school, normal-school, and university students in British Columbia on a modified Army Alpha Test showed differences when students were classified by paternal occupation; but median I.Q.'s did not vary so much from class to class as in many other similar studies. Six hundred fifty-nine subjects from professional classes had a median I.Q. of 105.1, and 456 from 'unskilled' classes had a median I.Q. of 100.8. The author thinks the lack of differentiation may be due to 'restriction' at the upper end of the scale.

Comment: Lack of differentiation is possibly due also to selection, since all subjects were in, or beyond, the high school.

56. Terman, L. M., and others. *Genetic Studies of Genius; Vol. 1, Mental and Physical Traits of a Thousand Gifted Children.* Stanford University, 1925. 648 pp.

The fathers of 560 children of the main group of the Stanford study of gifted children were distributed on the Taussig Scale in the following percentages:

Professional	31.4	Skilled labor	11.8
Semi-professional and business	50.0	Semi-skilled to slightly skilled	6.6
(a) Higher group	31.2	Common labor	0.13
(b) Lower group	18.8		

The author says: "Our data show that individuals of the various social classes present differences in early childhood—a fact which strongly suggests that the causal factor lies in original endowment rather than in environmental influences."

57. Visher, S. S. "A study of the type of the place of birth and of the occupation of fathers of subjects of sketches in '*Who's Who in America.*'" *Amer. Jour. Soc.*, 30: 1924-1925, 551-557. Also in *Who's Who*, 1924-1925.

The study is based upon persons in *Who's Who*, 1924-1925. "In proportion to population at the 1870 census, cities contributed nearly six

times as many as did farms; villages, nine times as many, and suburbs eleven times." Fathers who were in professional, business, farming, skilled or semi-skilled, or unskilled laboring classes contributed sons in the proportions 1400, 600, 70, 30, and 1.

The author comments: "This study does not indicate whether 'nurture' or 'nature' . . . is the more important."

58. Yerkes, R. M. (Editor) *Psychological Examining in the United States Army. National Academy Science Memoirs,* 15: 1921, Pt. 3, 890 pp. (Chapter 15. "Intellectual ratings of occupational groups.")

Data based upon average Alpha and Beta scores upon men from 114 occupations, in 16 camps, show pronounced decrease from the professional to laboring classes of occupations.

b. *Social or Cultural Status.*

59. Bickersteth, M. E. "The application of mental tests to children of various ages." *Brit. Jour. Psych.,* 9: 1919, 23-73.

City and country data. The conclusion is drawn that city children are slightly better on reasoning, country children slightly better on memory.

Comment: The possibility that racial differences or selective factors, rather than city or country differences, may account for the results gives rise to some ambiguity.

60. Bridges, J. W., and Coler, Lillian E. "The relation of intelligence to social status." *Psych. Rev.,* 24: 1917, 1-31.

Children in a 'favored' Columbus school score higher, age for age, on the Yerkes-Bridges Point Scale, than those in an 'unfavored' school. The authors admit that the results "could be used by adherents of the 'environment theory' as well as by advocates of 'inheritance.'"

61. Chapman, J. C., and Wiggins, D. M. "Relation of family size to intelligence of offspring and socio-economic status of family." *Ped. Seminary,* 32: 1925, 414-421.

For 632 children, including many foreigners, in Grades VI to VIII, Chapman-Sims social ratings and National Intelligence Test I.Q.'s were obtained. The correlation was $.32 \pm .02$.

Comment: The presence of many low testing foreigners may account in part for this correlation.

62. Decroly, O., et Degand, J. "La mesure de l'intelligence chez les enfants normaux d'après les tests de MM. Binet et Simon." *Archives de Psych.,* 9: 1910, 81-108.

Children of good social class in Belgium tested considerably above Binet's norms.

63. *Duff, J. F., and Thomson, G. H. "Social and geographical distribution of intelligence in Northumberland." *Brit. Jour. Psych.*, 14: 1923, 192-198.

About 13,000 children, representing practically all those over 11 and under 13 in certain elementary and secondary schools of Northumberland, were tested on the Northumberland Mental Test. The average I.Q.'s of pupils from various classes varied from 121 (children of clergymen) down to 91 (hawkers' and chimney sweeps' children) and 88 (children of insane and criminals.)

Results classified as to locality tend to bear out an earlier suggestion that children of the cities and *remote* rural sections test higher than those in ordinary rural sections, thus showing that the low average scores usually found for rural children may be due to selection rather than to environment.

64. Fukuda, T. "A survey of the intelligence and environment of school children." *Amer. Jour. Psych.*, 36: 1925, 124-139.

Stanford Binet Tests and environmental scores (based upon a combination of the Whittier Home and Neighborhood scales) were obtained for 257 elementary-school pupils. The median I.Q. was 93. The correlation between I.Q. and environmental score was .53.

Comment: Part of this correlation may be due to the presence of 129 foreign children.

65. *Galton, F. *Hereditary Genius.* London, 1869, 390 pp.

The author writes: "The practice of nepotism among ecclesiastics is universal. (Among Popes and other Roman Catholic dignitaries). It consists in their giving those social helps to a nephew, or other more distant relative, that ordinary people give to their children. . . . I do not profess to have worked up the kinships of the Italians with any especial care, but I have seen amply enough of them to justify me in saying that the individuals whose advancement has been due to nepotism are curiously undistinguished."

66. Gaw, Frances. "A study of performance tests." *Brit. Jour. Psych.*, 15: 1924-25, 374-392.

Twenty-seven canal-boat children between ages 6 and 13 were tested by the Stanford Binet and a battery of 14 performance tests. Binet I.Q. and performance I.Q. correlated .68. Educational ratios based upon Burt and Ballard tests correlated .26 with performance I.Q. and .58 with Binet I.Q.

Comment: The author states that the results "seem to justify the conclusion that the performance tests are less influenced by environ-

ment and by lack of schooling than is the Binet scale." She does not suggest the possibility that Binet may be merely a better index of capacity to learn.

67. Goodenough, Florence L. "The relation of the intelligence of pre-school children to the education of their parents." *School and Soc.*, 26: 1927, 54-56.

"A total of 213 children, all of whom were between the ages of 18 and 54 months, were given the Kuhlmann Binet test twice with an interval between the tests of approximately six weeks. . . . The correlation between the I.Q. earned on the first test and the education of the mother was $.32 \pm .04$; with the education of the father $.26 \pm .04$. On the second test, the correlation with the mother's education was $.35 \pm .04$; with the father's education $.35 \pm .04$. . . . Since the influence of heredity is presumably exerted equally by the two parents, the close agreement of the correlations suggests that the relatively large proportion of time which the child of pre-school age spends with the mother as compared with that spent with the father has little effect upon the outcome of these tests."

68. *Gordon, Hugh. *Mental and Scholastic Tests among Retarded Children.* (London Board of Education, Ed. Pamphlets, No. 44, 1923, 92 pp.)

See section on Intelligence and Schooling for summary.

69. Hirsch, N. D. M. "A summary of some of the results from an experimental study of the East Kentucky mountaineers." *Proc. Nat. Acad. Science*, 13: 1927, Pt. 1, 18-21.

The conclusion is drawn, in a way not made clear, that environment accounts for about 37 percent of the deficiency of the subjects.

70. Hoffman, A. "Vergleichende Intelligenzprüfungen und Vorschülern und Volkschülern." *Zsch. f. angew. Psych.*, 8: 1914, 102-120.

One hundred fifty-six boys from a *Vorschule* and a *Volkschule* were tested by Bobertag's Binet revision. Ten-year-old boys in the *Volkschule* averaged no better than nine-year-olds at the *Vorschule*.

71. Ide, Gladys G. "The increase of the intelligence quotient through training." *Psych. Clinic*, 14: 1922, 159-162.

Five girls are reported in a "private institution endowed for the education of fatherless girls" whose I.Q.'s rose after they entered the school. Two I.Q.'s rose as many as 16 points.

Comment: Precise evaluation is impossible because of the lack of any control group, and the lack of any statement as to how many I.Q.'s *failed* to rise.

72. Isserlis, L. "The relation between home conditions and the intelligence of school children." *Medical Research Council Special Report Series*, No. 74, 1923.

From results of group tests and of teachers' ratings of intelligence, made upon several hundred children, which correlated with environmental measures, such as social status, condition of clothing, and care of home, about .30 to .35, the author concludes that "progressive improvement in home conditions may be expected to react favorably not only on the health, but also on the intelligence of school children." Burt, in his preface to the work, says that the author "demonstrates that the influence of home conditions upon intelligence is a positive one." But a few paragraphs later he says it "remains impossible to decide how far the child from the better-class home is profiting by mere environment. . . . and how far he is profiting from heredity, simply inheriting the high intelligence which had already guided his parents or his ancestors into a superior class." (See critical notice by Pearson, No. 77.)

73. *Jones, D. C., and Carr-Saunders, A. M. "The relation between intelligence and social status among orphan children." *Brit. Jour. Psych.*, 17: 1927, 343-364.

The Simplex Group Test (an English group mental test) was given to children and their I.Q.'s were computed. The pupils were 9 to 15 years old. A total of 700 or 800, from eight industrial schools or orphanages, appeared in some of the treatments. The basal age used is not stated. The social origin of the children was rated on a scale very similar to the five-category Taussig Scale, on which Class 1 is highest.

For pupils 9½ to 14 in all the schools, the following mean I.Q.'s were found:

1. Pupils in residence from 0 to 3 years
   - Occupation Class 1 and 2.................. 107.5
   - Occupation Class 3......................... 97
   - Occupation Class 4 and 5.................. 94.5
2. Pupils in residence over 3 years
   - Occupation Class 1 and 2.................. 106.5
   - Occupation Class 3......................... 98.5
   - Occupation Class 4 and 5.................. 96.5

Thus, the differences tend to decrease slightly with length of residence, 0-3 years and over 3 years.

Difference in Mean I. Q. According to Length of Residence

I. Q. Greater In Earlier Period	Diff.	P.E.	I. Q. Greater In Later Period	Diff.	P.E.
Occupation 1 and 2......	9.7	3.7	Occupation 3..........	1.0	2.4
Occupation 1............	4.3	1.8	Occupation 4 and 5.....	0.6	1.7
Occupation 2............	5.9	3.0	Occupation 4 and 5.....	0.3	1.7
Occupation 4 and 5......	0.7	2.5	Occupation 4 and 5.....	2.1	1.8

The pupils are classified by schools and seem to show a differential effect of environment on different social classes.

The authors conclude: "There is no reason to suppose, so far as this evidence is concerned, that environmental influences are the whole or even the major part of the cause of the differences in intelligence between children of different social origin."

Comment: The result of this study may be compared with interest with those of the Chicago and the Stanford studies of foster children appearing in this Yearbook. Environmental effects were found in these two studies of about the same magnitude as those in the present study (6 or 7 points). The possibility of chronological selective factors, which in orphanage studies of the past has not been ruled out, seems not to cloud seriously the results of this study, since the average I.Q. of one class of children is seen to *increase* with length of residence, and of another class of children, to *decrease*.

74. Kornhauser, A. W. "The economic standing of parents and the intelligence of their children." *Jour. Educ. Psych.*, 9: 1918, 159-164.

Of the families of 1000 Pittsburgh school children, 32 percent had telephones. In families of children retarded in school one half year or more, 19 percent had telephones; 32 percent of families of children normally advanced, and 50 percent of families of accelerated children had them.

75. Morlé, M. "L'influence de l'état social sur le degré de l'intelligence des enfants." *Bull. Soc. Libre Educ. Psych. de l'Enfant*, 12: 1911, 8-15.

Thirty random children from a 'favored' school and 30 random children from an 'unfavored' school were compared. The children from the poorer section were, on the average, one-fourth year behind the level of their age. Those of the 'favored' school averaged from one-fourth to one-half year accelerated.

76. Odin, Alfred. *Genèse des Grands Hommes. Gens de Lettres Français Modernes.* 2 vols., Paris, 1895.

French men of letters were relatively more often born in cities than in the country. The author believes that this fact indicates an environmental influence upon literary achievement.

77. Pearson, K. "Critical notice: on the relation between home conditions and the intelligence of school children." *Biometrika*, 15: 1923, 161-172.

Pearson recomputes most of Isserlis' data, getting results which do not agree perfectly with his. Isserlis is criticized severely for (1) not

telling how his etas and contingency coefficients were computed; (2) not knowing from the literature that Heron and Gilby had obtained similar results before; (3) drawing conclusions when numbers were too low and P.E.'s too high; (4) partialling out variables when r's are not based upon the same cases; (5) concluding that "progressive improvement in home conditions may be expected to react favorably not only on the health but also on the intelligence of school children." He also criticizes Burt for claiming in the preface that home conditions and intelligence correlate between .3 and .4 when "all irrelevant factors have been eliminated." Only C. A. was partialled out. (See also the summary of the study of Isserlis, No. 71.)

78. Pintner, R. *The Mental Survey.* New York, 1918, 116 pp.

Groups of city school children averaged uniformly above the average level of village and rural school children.

79. Pressey, S. L., and Thomas, J. B. "A study of country children in (1) a good and (2) a poor farming district, by means of a group scale of intelligence." *Jour. Appl. Psych.*, 3:1919, 283-286.

In a study of 2800 city children and 538 country children, proportions varying from 21 to 33 percent of country children equalled or exceeded the median of city children.

80. Root, W. T. "The intelligence quotient from two viewpoints." *Jour. Appl. Psych.*, 6:1922, 267-275.

The author tables tests and retests of six children taken from inferior homes and placed in "a very superior institutional home on the cottage plan." Changes in I.Q. (after 12 to 19 months) ranged from — 9 to + 14, with an average of 5.5 points.

81. *Schmitt, M. "Der Einfluss des Milieus und anderer Factoren auf das Intelligenzalter." *Fortschr. d. Psych.*, 5:1919, 217-255.

One hundred children, aged 4 to 16, in two Würzburg orphanages were tested by a translation of Binet's 1911 Scale. The M. A.'s of the older children were corrected for the inadequacy of the scale at the upper end. All the children had come from an originally poor home environment.

When the subjects were divided into groups each covering a range of two years C. A., no significant difference was found in these groups between the proportion of mental retardation of those who had been in the institutions longest and of those who had been there shortest, or of those who had come from a poor moral environment and of those who had come from a better moral environment, or of those whose fathers

had been industrial workers and of those whose fathers had been peasants. A slight advantage was found in favor of those coming from the less abject material surroundings, but the author points out that this fact merely corroborates previous studies on economic status and proves nothing.

There is no consistent tendency for children within the various age groups who have been in the institutions longest to show less mental variability than those who have been there for a shorter time.

Comment: The author admits that the small numbers prevent results from being conclusive, but believes that any "large influences" would nevertheless be revealed if present.

Although the results appear to be negative with respect to environmental influence, there is the possibility that some type of selective influence may have covered up an existing environmental influence.

82. Strong, A. C. "Three hundred fifty white and colored children measured by the Binet-Simon Measuring Scale of Intelligence; a comparative study." *Ped. Seminary,* 20: 1913, 485-513.

So far as the white children were concerned, less than six percent of the city school children were retarded, while 18 percent of mill-district children were over a year below the mental level of their age. None of the mill-district children was mentally accelerated, although 10 per cent of city school children scored above average.

83. Terman, L. M. *The Stanford Revision and Extension of the Binet-Simon Scale.* Baltimore, 1917, 179 pp.

The I.Q.'s of 492 children out of the original Stanford Binet standardization group correlated with social status (on a five-point scale) to the extent of .40. Median I.Q.'s for children from five social levels were:

Social Group	I.Q.
Very inferior	85
Inferior	93
Average	99.5
Superior	107
Very superior	106

84. *Terman, L. M. *The Intelligence of School Children.* Boston, 1919, 313 pp.

Some individual cases are cited of (a) children showing good intelligence in poor environment, and (b) children showing no improvement when placed in a good environment. A child of seven who had never been to school, though he had excellent home advantages, tested at 153.

Gypsy Mary, aged 16, stolen by gypsies when she was about four, had had no schooling. She tested at approximately 100, and in a year after being freed was ready for the high school.

Two boys were adopted at ages five and seven from a very poor home. Their I.Q.'s were 73 and 72. Four years later, after exceptional advantages, their I.Q.'s were only 70 and 77.

85. Terman, L. M., and Wagner, Dorothy. "Intelligence quotients of 68 children in a California orphanage." *Jour. of Delinquency*, 3:1918, 115-121.

Children, originally from poor homes, who had lived in an orphanage anywhere from a few days to 11 years had a median I.Q. of 92.3. "The correlation between intelligence quotient and time in the institution . . . was found to be slightly negative (—.16). This gives no support to the environment hypothesis."

Comment: Since there would probably tend to be a positive correlation between age and length of residence at the orphanage, and since the Stanford Binet I.Q. shows a slight negative correlation with age, the slight negative correlation of I.Q. with residence might perhaps be expected.

The lack of environmental influence cannot be considered as thoroughly established, since selective influence upon children of various ages might conceivably be sufficient to cover up the influence of environment if it existed.

86. *Theis, Sophie V. S. *How Foster Children Turn Out.* N. Y., 1924. (State Charities Aid Ass'n., Publ. No. 165.)

Data were collected by 16 trained field workers upon 910 former wards of the State Charities Aid, 18 years old or over by 1922. Of these, 83 percent were under 25. Thirty-five percent were under five when placed, 21 percent between five and 10, and 45 percent between 10 and 19. It was possible to secure considerable information upon 797 of the 910.

For each of 797 subjects, information was secured on "capability to manage himself and his affairs with ordinary prudence." Of the total group, 77 percent were 'capable' and 23 percent were 'not capable.'

When 604 subjects who had had only one kind of home since placement (superior, good, or mediocre with respect to cultural and material status) were compared for the proportion 'capable' from each type of home, no home influence was manifest. Negative results were also found when the proportions of 'A,' 'B,' and 'C' subjects, in ability to take formal education, were tabulated against type of foster home. ('A' ability was defined as "capable of progressing beyond the eighth grade" or "having an intelligence quotient of 80 or more.")

Of 754 children who had had only one kind of care since placement, 12 percent had had excellent care, 68 percent average care, and

20 percent poor care (by the standards adopted for the study). Of those "who had excellent care, approximately 87 percent developed into 'capable' subjects; of those who had average care 80 percent are 'capable,' and of those who had poor care 66 percent are 'capable.' In other words, when excellent care was given, more of the subjects developed into capable persons."

When the type of family background from which the subjects had originally come was tabulated, it was found that, in a group of 693 subjects, 435 had predominantly 'bad' family background; 65 had a mixed 'good-and-bad' background; 44 had a 'good' background, and 149 an 'unknown' background. The percentages of 'capable' subjects in these four classes were 71, 80, 83, and 81. This comparison seems to favor children with better background, but according to the author the differences here reported are not statistically reliable. Corresponding percentages of "A" subjects in ability to take formal education were 63, 67, 92, and 77. The differences in this case are regarded as significant.

Of 271 subjects placed under the age of five whose 'capability' was known, 86 percent were capable. Only 73 percent of children placed at five or over were capable. Ninety percent placed under five years of age were rated 'A' on ability to take education as against 65 percent of those placed at five or over.

Of 155 subjects one or both of whose parents had been mentally inferior, 68 percent were capable, and 53 percent were rated "A" on ability to take education. Of 82 subjects known to have had good ability in their parentage, 80 percent were capable, and 72 percent were rated 'A.'

Comment: No unjustifiable conclusions appear to be drawn from the data, which, as the author admits, lack precision. Upon the statistical, as well as upon the clinical, data, the author bases the statement that "undoubtedly the child's adjustment to his foster family governs to a significant degree his adjustment to society, and his adjustment to his foster family has less to do with their standards of comfort and their place in the community than with their human qualities and their understanding." The results as regards 'capability' and 'A' ratings seem to favor heredity somewhat more than environment, but, as the introduction by Homer Folks points out, "We cannot disentangle the factor of inheritance from that of early life with the children's parents and the environment provided by them."

87. *Thomson, G. H. "The Northumberland Mental Tests." *Brit. Jour. Psych.*, 12: 1921, 201-222.

The pertinent part of the study concerns 64 pupils in small schools in the Cheviot valleys (very remote), two schools in a poor suburb of Newcastle involving 64 pupils, and 111 pupils in two schools in a well-to-do suburb of Newcastle. The Northumberland Mental Tests (whose

norms were based upon 2532 cases) were given chiefly for the purpose of discovering scholarship candidates in rural districts.

In trials with 300 cases, children from a "very rural district" tested about a year behind children "from a large city" (the age of these pupils is not stated). But in three schools in the Cheviot valleys (where little drainage of descendants of Border troopers to the cities had taken place), 80 percent of 25 pupils tested above 100 I.Q. in one school, 70 percent of 17 in another, and 50 percent of 22 in another. These percents were higher than in suburban regions. In a "poor suburb," percents above 100 I.Q. were 29 and 35; in a well-to-do suburb, 55 and 56. The study corroborates Miss Bickersteth's results—*i.e.*, ability is highest *close to* and *far away* from cities.

Comment: These results suggest that the sectional differences found so universally upon intelligence tests may be due much more to selection and much less to environment than the advocates of environment believe.

88. Yerkes, R., and Anderson, Helen. "The importance of social status as indicated by the results of the point-scale method of measuring mental capacity." *Jour. Educ. Psych.*, 6: 1915, 137-150.

Kindergarten and first-grade children in a well located and in a poorly located school were tested. Pupils in the well located school averaged much the higher, except in the four-year-old group. Four-year-olds in a poorly located school possibly averaged higher because they were less timid.

3. Race Differences

a. *Differences Between Racial Groups.* These differences are important to know, but almost no studies of this type provide any means for evaluating the contributions of nature and nurture. The environment of low-testing races or nationalities is generally inferior to that of high-testing ones, but we do not know whether this environmental difference is a cause or an effect. Until we can have controlled experiments upon children of various races *transplanted at infancy* into uniform environments, precise knowledge regarding native racial differences may be impossible to secure.

89. Arlitt, A. H. "On the need of caution in establishing race norms." *Jour. Appl. Psych.*, 5: 1921, 179-183.

Because differences on the Stanford Binet between children of high and of low economic status were greater than differences between average levels of American whites, Italians, and Negroes, the author urges

that valid racial comparisons can be made only between groups of equal economic status. The median I.Q. of whites was 106; of Italians, 85; and of Negroes, 83.

Comment: The author fails to consider that the different intelligence levels for groups of varying economic status may themselves be congenital, *i.e.*, that individuals of certain mental levels may be drawn to certain occupations, rather than that economic status conditions intelligence. To make racial comparisons only between groups of equal economic status would introduce inextricable factors of selection.

90. Boas, F. *The Mind of Primitive Man.* New York, 1922, 294 pp.

Chapter I of this book has been summarized in order to illustrate the point of view of one school of thinkers with respect to mental differences in various races. Boas protests against the frequent, but "unproved" assumption that the mental status of a people is measured by the difference between its social status and our own. He contends, in the first place, that the fact that certain racial groups had developed a few thousand years ago civilizations superior to that of our own ancestors is not evidence that our race was then an inferior one. No more is the white superiority over the achievements of many of the so-called 'primitive races' of this era evidence of a higher endowment of the white race. The dissemination of culture is a slow and partly accidental process, and the arrival of one race at a certain point of culture a thousand or two years earlier than another race may be a matter of chance.

Since real contributions to civilization are made by relatively few members of a race, and since a great share of civilized growth takes place through the spread of culture from other groups, the problem resolves into one of *assimilation* of culture. Here, obviously, many 'primitive' races are weak. The author, citing a few examples of the following conditions, but without proving that they are actually responsible for the inferiority of the races concerned, concludes that the question of racial differences is still unsettled. The conditions that may account for the slow progress of some races are: (1) the difference in appearance of primitive and civilized races, and consequent racial prejudices; (2) the ravages of diseases, which have usually been introduced to primitive people by white settlers; (3) the fact that hand-made products cannot compete with manufactured ones brought in with white settlers in great abundance, so that the progress of 'home industry' is blocked; and (4) the circumstance that the primitive people are sometimes crowded out of their accustomed haunts by sheer force of numbers.

Comment: While the study has little that could be called evidence either for or against innate racial difference, it presents speculations of an important nature.

91. Brigham, C. C. *A Study of American Intelligence.* Princeton, 1923, 210 pp.

This is a further analysis of data on the American army intelligence tests. The chief contribution, other than recapitulation of results reported in Volume 15, National Academy of Science Memoirs, is the tabulation of test scores by the estimated proportions of Nordic, Alpine, and Mediterranean racial strains in men from the European countries. Weighting the score distributions of the foreign countries in proportion to the estimated number of each racial strain comprising the foreign country in question, an "approximate sampling" was obtained to represent the level of the three races. These scores[2], with the S. D.'s of the distributions were:

	Nordic	Alpine	Mediterranean
Mean	13.28	11.67	11.43
S. D.	2.70	2.87	2.70
N	3456	4766	4196

When non-English speaking Nordics were used in similar comparisons, differences persisted, but became smaller.

The average scores of men who had been in this country for varying lengths of time were:

0-5 yrs.	6-10 yrs.	11-15 yrs.	16-20 yrs.	Over 20 yrs.
11.41	11.74	12.47	13.55	13.82

These differences probably could not be entirely racial ones, because the period during which the greatest drop in proportion of Nordic immigration to this country occurred does not correspond to an especially marked change in average score for the above classified groups. Differences are perhaps accounted for in part by increasingly inferior immigrants and in part by the effect of Americanization over a period of years.

92. Brown, G. L. "Intelligence as related to nationality." *Jour. Educ. Research,* 5: 1922, 324-327.

Stanford Binet tests given to 913 school children yielded median I.Q.'s that ranged from 104 for Norwegians, 102 for Swedes, Germans, and English, down to 86 for Slovaks, and 78 for Italians.

93. Clark, W. W. "Birth rate and native intelligence." *Psych. Clinic,* 14: 1922, 111-115.

Los Angeles negro children in five schools did practically as well on the National Intelligence Test and a battery of achievement tests

[2] The scores, whether on Alpha, Beta, or Stanford Binet, were transmuted into comparable units.

as did white children. The children were from 15 representative Los Angeles schools.

Comment: Selection probably accounts for this result, since no other studies have yielded equivalent performance for negro and white groups.

94. Darsie, M. *The Mental Capacity of American-Born Japanese Children.* Baltimore, 1926, 89 pp. (Comp. Psych. Monograph, 3, No. 15.)

Five hundred seventy urban Japanese children (representing practically all children from 10 to 13 in certain communities) tested from 6 to 12 points below white children of those ages on the Stanford Binet, but on the Army Beta Test the Japanese were equal to the white at age 10 and improved, relative to the whites, until 75 percent exceeded the white median at age 12.

The author suggests that the discrepancy on Beta at different ages may be explained by the fact that Beta is not difficult enough to distribute ability in American children after age 11; but in the case of the Japanese, 82 percent of their superiority after age 12 is due to performance on the digit-symbol and number-comparison tests, which provided ample range. The author thinks that this may represent an "industry" factor, or factor of "sustained attention."

On the Stanford Achievement Test the averages of total Japanese scores from age 10 to 13 were approximately at the white norm, but on reading, language, and science, the Japanese at age 13 fell behind the norm almost a year.

Seven judges were asked to rate Binet items for their dependency upon linguistic skill. The rank-order correlation composite judgments of linguistic dependency and the degree of superiority of white over Japanese on specific items was .62 or .65, depending on how the superiority was measured. The author concludes: (a) "There is a very definite tendency for the Japanese to equal or exceed American norms as they pass from the linguistic to the non-linguistic tests." (b) "It is highly probable that the innate mental capacity of Japanese children is greater than the Binet I.Q. would indicate. How much greater the writer is unable to estimate." (c) "Japanese superiority in the enclosed boxes, Binet paper cutting, and code tests is established beyond question." These, like the digit-symbol and number-comparison tests of Beta, seem to require acuity of visual perception and sustained attention.

95. Fukuda, T. "Some data on the intelligence of Japanese children." *Amer. Jour. Psych.*, 34:1923, 599-601.

Forty-three Japanese children in Denver had an average I.Q. of 97.

96. Garth, T. R., and Whatley, C. A. "The intelligence of Southern negro children." *School and Soc.*, 22:1925, 501-504.

The average I.Q. on the National Intelligence Test of 1300 negro children in southern cities was 75.

97. Goodenough, Florence L. "Racial differences in the intelligence of school children." *Jour. Exper. Psych.*, 9: 1926, 388-397.

The subjects were "2457 public-school children (in Grades I to IV) practically all of whom were American born, but in whose immediate ancestry a number of racial stocks are represented. Children of mixed parentage have not been included, except in the case of the Negroes . . . and the Indians."

The Goodenough Intelligence Test for Young Children (drawing a man) was applied. Means, medians, S. D.'s and coefficients of variability were computed for 16 groups.

Mean I.Q.'s varied from 106.1 (Jewish); 104.1 (Chinese); 103.5 (Scandinavian); 101.9 (Japanese); 101.5 (American); down to 89.1 (Italian); 88.5 (Spanish-Mexican); 85.8 (Calif. Negroes); 85.6 (Hoopa Valley Indians); 78.7 (Southern Negroes). The results from this non-verbal test thus seem to agree with other results from verbal tests.

This article includes an excellent tabular summary of representative studies on test scores of school children as conditioned by race or nationality. The summary lays no claim to being complete, however, and omits a number of important groups, such as Russian and Japanese. The summary is reproduced on p. 287 because it presents many data in compact form.

98. Kubo, Y. "The revised and extended Binet-Simon tests applied to the Japanese children." *Ped. Seminary*, 19:1921, 187-194.

An adaptation and translation of the Binet Scale given to 1200 Japanese children of ages two to 15 in Tokyo yielded average I.Q.'s of 98 and 99 at various ages.

99. Mayo, M. J. *The Mental Capacity of the American Negro.* New York, 1913, 70 pp. (Archives of Psych., No. 28.)

Of 150 negro high-school pupils under economic conditions similar to those of 150 white high-school pupils, 29 percent of negroes exceeded the white median of high-school marks.

100. Murdoch, Katherine. "A study of race differences in New York City." *School and Soc.*, 2: 1920, 147-150.

Hebrew, American Negro, and Italian children of approximately the same economic status and all known to have no language handicap were given the Pressey group tests. The median of the Hebrews was surpassed by 57.7 percent of Americans, 30.3 percent of Negroes, and 15.5 percent of Italians.

INTELLECTUAL ABILITY OF AMERICAN SCHOOL CHILDREN OF VARIOUS RACIAL STOCKS (After Goodenough)

Racial Stock	Reported by	Number of Cases	Test Used	Result in I.Q. (Unless otherwise stated)
White American.......	Pintner and Keller......	249	Stanford-Binet	95
White American.......	Sheldon......	100	S. B.	104
White American.......	Dickson......	49	S. B.	106
American Negro (Ohio)	Pintner and Keller......	71	S. B.	88
American Negro....... (Tennessee)	Peterson......	All children 8-10 yrs. in several schools.....	Pressey......	75
American Negro....... (Arkansas)	Jordan.......	247	N. I. T.......	14 yr. negroes equal 10 yr. whites
American Negro....... (Northern)	Thorndike....	349	I. E. R.......	Ca. 4% negroes passed median score for whites of same grade
American Negro.......	Sunne........	Over 1000....	N.I.T. and Myers......	Negro ave. 1-1½ yrs. M.A. below whites
English..............	Pintner and Keller......	24	S.B...........	97
English..............	Brown.......	90	S.B	101.8
Italian..............	Pintner and Keller......	313	S.B...........	84
Italian..............	Dickson......	25	S.B...........	84
Italian..............	Young.......	Several hundred.......	Army Alpha and Beta...	About 83
Italian..............	Brown.......	51	S.B...........	77.5
German..............	Pintner and Keller......	37	S.B...........	91
German..............	Brown.......	67	S.B...........	102.3
Jewish..............	Pintner and Keller......	79	S.B...........	95
Jewish..............	Murdock.....	Several hundred.......	Pressey.......	Jews approx. equal to white Americans
Chinese (San Francisco)	Yeung.......	109	S.B...........	97
Chinese (Hawaii)......	Symonds.....	513	Pintner non-language...	99
Spanish-Mexican.......	Sheldon......	100	S.B. and Cole-Vincent....	89
Spanish-Mexican.......	Dickson......	37	S.B...........	78
Portuguese...........	Young (Quoting unpub. data by Roll)	119	S.B...........	86
Portuguese...........	Dickson......	23	S.B...........	84
Norwegian...........	Brown.......	34	S.B...........	103.8
Swedish.............	Brown.......	187	S.B...........	101.9
Austrian.............	Brown.......	28	S.B...........	99.5
French..............	Brown.......	199	S.B...........	95.4
Finnish..............	Brown.......	226	S.B...........	90
Slavish..............	Pintner and Keller......	130	S.B...........	85
Hungarian...........	Pintner and Keller......	99	S.B...........	89
Indian (Michigan)....	Rowe........	268	Goddard-Binet	Only 5.8% of Indians tested at age or above.
Indian..............	Hunter.......	711	Otis..........	Whites excel Indians by 1.6 P.E. of latter. Correlation bet. degree of white blood and score = .51 ± .017.

101. Peterson, Joseph. *The Comparative Abilities of White and Negro Children.* Baltimore, 1923, 141 pp. (Comp. Psych. Monograph, 1: No. 5.)

Three hundred fourteen negro children, aged 7 to 10, were about 1.5 P. E. units inferior to white children, both upon Peterson's Rational Learning Test and upon elementary, verbal group tests. Two hundred eighty-four of them, roughly divided into five classes on the basis of color, yielded a mean square contingency coefficient between color and Rational Learning score of .50.

Comment: It is significant to know that differences exist even on tests which would appear to be little influenced by schooling, although this study, like most other studies of racial difference, does not separate the influence of heredity from that of environment in any crucial way.

102. Pyle, W. H. "A study of the mental and physical characteristics of the Chinese." *School and Soc.*, 8: 1918, 264-269.

Children in China tested on a battery of five mental tests show a superiority to American children on rote memory, but an inferiority to Americans on logical memory, substitution, analogies, and spot pattern.

103. Schwegler, R. A., and Winn, Edith. "Comparative study of intelligence of white and colored children." *Jour. Educ. Research*, 2: 1920, 838-848.

Fifty-eight white boys and girls in a Kansas Junior High School tested 10 points higher on the Stanford Binet than 58 negro boys and girls.

Comment: The negro children were probably somewhat selected, owing to attendance at junior high school. Yet even so, a significant racial difference was present.

104. Seago, D. W., and Koldin, T. S. "A comparative study of the mental capacity of sixth-grade Jewish and Italian children." *School and Soc.*, 22: 1925, 564-568.

Jewish boys test considerably higher than Italian boys on the National Intelligence Test.

105. Strong, A. C. "Three hundred fifty white and colored children measured by the Binet-Simon Measuring Scale of Intelligence." *Ped. Seminary*, 20: 1913, 485-512.

Of children tested in Columbia, S. C., schools, 60.8 percent colored and 25.2 white were below the norm, while 9.2 percent colored and 26 percent white were above the norm. (Cf. abstract of same article, No. 81.)

106. Yeung, K. T. "Intelligence of Chinese children in San Francisco." *Jour. Appl. Psych.*, 5: 1921, 267-274.

One hundred nine Chinese children in San Francisco had a median I.Q. of 97 on the Stanford Binet. Tests were translated into Chinese when marked language handicap was present.

107. Yerkes, R. M. (Editor) *Psychological Examining in the United States Army. National Academy of Science Memoirs,* 15: 1921, 890 pp.

Chapter 6: "Relation of intelligence ratings to nativity."

Some 13,000 out of 94,000 of the white draft of the principal sampling were foreign born. Only countries from which 100 or more cases were reported are considered.

The rank order of countries on the basis of men obtaining A or B ratings (apparently on whatever test they took—Alpha, Beta, Binet, or Performance) is that shown in the accompanying table.

Country	Percent	Country	Percent
England	19.7	Ireland	4.1
Scotland	13.0	All foreign	4.0
White draft	12.1	Turkey	3.4
Holland	10.7	Austria	3.4
Canada	10.5	Russia	2.7
Germany	8.3	Greece	2.1
Denmark	5.4	Italy	0.8
Sweden	4.3	Belgium	0.8
Norway	4.1	Poland	0.5

Tables are also given of total Alpha and Beta distributions, total letter-grade distributions, etc.

Chapter 11: "Intelligence of the draft."

In a sampling from all states, 19,000 negroes had a median M. A. of 10.1.

Chapter 10: "Statistics on education." Three thousand negroes had a median Alpha score of 38.6, and 2000 southern negroes a median of 12.4. These values are contrasted with scores for 13,000 officers, 139.2; 52,000 native-born whites, 58.9; and 4000 foreign-born whites, 46.7.

108. Young, K. "Mental differences in certain immigrant groups." *University of Oregon Publications,* 1: 1922, No. 11. Also *Scient. Mo.,* 15: 1922, 417-434.

Italian, Portuguese, and Mexican 12-year-olds from California tested in the neighborhood of one S. D. below American 12-year-olds on

both Alpha and Beta test. (See also data from Yeung in the tabled summary of Goodenough in this section.)

b. *Differences Between Groups of Mixed Blood.* In so far as the environment of individuals of mixed blood is similar to that of full-blooded individuals, evidence of this kind is important. But the degree to which environment is similar is extremely problematic.

109. Ferguson, G. O., Jr. *The Psychology of the Negro.* New York, 1916, 138 pp. (Archives of Psych., 5, No. 36.)

Three hundred nineteen negro pupils in Richmond were divided on the basis of skin color into full, three-fourths, one-half, and one-fourth blood mixtures. As the author states, whatever inaccuracy enters into the classification would restrict, rather than emphasize, differences. On three tests the average percent of negroes aged 12 to 17 reaching or exceeding the white norms was that shown here in tabular form:

	Full and Three-Fourths	One-Half and One-Fourth
Test I, Mixed relations		
Boys	17.0	38.5
Girls	25.5	41.3
Test II, Mixed relations		
Boys	19.2	46.8
Girls	27.2	42.8
Completion Test		
Boys	19.0	44.8
Girls	23.2	31.5

The author points out that these comparisons on the basis of amount of negro blood would be even more striking if selection in the schools did not weed out negroes so rapidly. He thinks the results are strong evidence for heredity, since the social limitations of mixed-blood negroes are about the same as those of full-bloods.

Comment: Though possible differences in social status are not altogether ruled out, the results are most suggestive.

110. Garth, Thomas R. "A comparison of the intelligence of Mexican and mixed- and full-blood Indian children." *Psych. Rev.,* 30:1923, 388-401.

The National Intelligence Test was administered to 307 Mexicans, 126 mixed-blood Indians, 176 Plains Indians and South-East full-bloods, 249 Pueblo full-bloods, and 85 Navajo and Apache full-bloods. The In-

dians were tested in Indian schools at Chilocco and at Albuquerque. The Mexicans were tested in San Antonio schools.

Groups were subdivided by ages from 12 to 19. In each sub-group the order of merit was: mixed-bloods, Mexicans, Plains and South-East, Pueblo, Navajo and Apache. The mixed-bloods had 80 percent exceeding the median of the Plains and South-east at ages 12 and 13, and 60 percent at ages 18 and 19. Estimates of social status indicated the same sequence as the foregoing, and the average amount of education of the blood groups followed the same sequence, except that the Pueblo Indians had slightly more education than the Plains and South-East Indians. Because of the differences just mentioned, the factor of environment cannot be untangled.

111. Hunter, W. S., and Sommermier, Eloise. "The relation of the degree of Indian blood to score on the Otis Intelligence Test." *Jour. Comp. Psych.*, 2: 1922, 257-277.

Seven hundred eleven Indian pupils (chiefly between 14 and 20) were divided into four classes on the basis of degree of Indian blood. The correlation between degree of Indian blood and Otis Group Test scores was $.35 \pm .02$. This became .41 when age and months of schooling were partialled out. Schooling and Otis correlated .35 for the entire group, and the correlation varied only between about .30 and .40 when "degree of blood" groups were taken separately, from which the conclusion was drawn that schooling is not a prominent factor in the correlation between degree of blood and Otis.

The authors admit their material is limited by the fact that no data are available on the relative social status of the mixed groups. Also, the possibility of temperament accounting for differences is not ruled out, though observation of "the manner in which the students worked upon the tests," and the fact that all were accustomed to school work, lead the authors to infer a racial difference.

112. Peterson, Joseph. *The Comparative Abilities of White and Negro Children*. Baltimore, 1923, 141 pp. (Comp. Psych. Monograph, 1: No. 5.)

See summary in Part *a* of the section on race differences, No. 101.

113. Reuter, E. B. "The superiority of the mulatto." *Amer. Jour. Sociol.*, 23: 1917-1918, 83-106.

The author argues upon evidence that while most outstanding negroes are individuals of mixed blood, it is not known to what degree this fact is due to racial superiority of the white race, and to what degree to the "social superiority" of the mulattoes operating "to bring into the group the favorable variant types produced by the entire race."

c. *Attempts to Evaluate the Effects of Foreign Language Handicap.* Such attempts are usually made by comparing the

results of verbal and non-verbal tests; but, as is pointed out in Chapter II on "Statistical Hazards," verbal and non-verbal tests are not comparable. Correlations between length of time English has been spoken and performance on verbal tests provide the most useful approach.

114. *Bere, May. *A Comparative Study of the Mental Capacity of Children of Foreign Parentage.* New York, 1924, 105 pp. (Teachers College, Columbia University, Contrib. to Ed., No. 154.)

In groups of Hebrew, Bohemian, and Italian children, each comprising 100 ten-year-old boys, the following results were secured with a verbal and a non-verbal test:

	Stanford Binet I.Q.		Stanford Binet M.A.		Pintner-Paterson M.A.	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Italian.............	85	10	108	12	109	21
Bohemian..........	93	10	118	15	123	16
Hebrew............	98	13	123	16	102	20

The P.E. of the difference in M.A.'s between the two tests for each nationality or between any two nationalities on either test is about one month.

To test out the theory that the diminution of differences between the three groups on the Pintner-Paterson is due to elimination of language handicap, comparison was made of the average length of residence in this country of parents of total groups and of parents of children having higher rating on the Pintner-Paterson than on the Binet. Thus:

Italian, 19.47 years versus 19.3 for total group;
Bohemian, 19.21 years versus 19.7 for total group;
Hebrew, 18.53 years versus 17.56 for total group.

It was found also that among the children having a higher rating on the Pintner-Paterson, foreign language exclusively was used in the homes of:

23 percent of the Italians versus 23 percent of all the Italians;
19 percent of the Bohemians versus 29 percent of all the Bohemians;
0 percent of the Hebrews versus 3 percent of all the Hebrews.

The author concludes that: "The zero correlation between Stanford Binet and Pintner-Paterson Performance Test results and average length of residence, plus the rise in Stanford Binet and Perform-

ance mental ages with increased use of English in the home . . . do not support the argument that a language handicap is responsible for low I.Q.'s on the Stanford Binet tests."

The differences between results on verbal and non-verbal tests "suggest . . . . that these groups may differ in the nature, as well as in the amount, of their abilities."

115. Colvin, S. S., and Allen, R. D. "Mental tests and linguistic ability." *Jour. Educ. Psych.*, 14: 1923, 1-20.

Fifty children of American parentage and 50 of Italian parentage in the Providence schools were tested on the Stanford Binet, the National Intelligence Test, and the Lippincott-Chapman Classroom Products Survey Tests. They ranged in age from 11 to 16, with an average C.A. of 13.1 in each case. All were in Grades V to VIII. How they were selected is not stated. The results are shown in the table.

	American	Italian
Binet I.Q.	92	91
N.I.T. I.Q.	85	76
Pedag. age (presumably on Lippincott-Chapman)	11.8	12.0

Since the Americans and Italians had equal standing on the Binet and on Achievement, the authors attribute the 9 points difference on N.I.T. to language handicap. Binet and N.I.T. correlated .73 for the Americans and .79 for the Italians.

Comment: The conclusion seems reasonable, though not definitely proved, because Binet and N.I.T. measure abilities which do not entirely overlap (correlating with each other only about .70 to .80 in single school grades).

116. Feingold, G. A. "Intelligence of the first generation of immigrant groups." *Jour. Educ. Psych.*, 15: 1924, 65-82.

American children of foreign parentage in a Hartford high school were tested by Army Alpha. The rank order of average scores for the different nationalities was almost identical with that of the army results. Considering here only the results from the freshman class, the largest difference between groups in mental age was 11 months (between British and Polish, and British and Italian). This, compared with the mental age difference of two years between native and foreign-born draftees, suggests that some of the army differences may have been due to language handicap.

Comment: The comparative uniformity of the high-school pupils may have been due to selection exercised by attendance at the high school.

117. *Madsen, I. N. "Some results with the Stanford Revision of the Binet-Simon Tests." *School and Soc.*, 19: 1924, 559-562.

Sixteen Italian children thought by their teachers to have language handicaps were tested on entering the first grade and retested a year and a half later. The mean I.Q. on the first test was 79.3, and on the second test, 80.1. The correlation between test and retest was .97 ± .01. No evidence appears of any influence of language handicap.

118. Pintner, R. "Comparison of American and foreign children on intelligence tests." *Jour. Educ. Psych.*, 14: 1923, 292-295.

American third- and fourth-grade pupils show less superiority to foreign pupils on the Pintner Non-Language Test than on the National Intelligence Tests.

Comment: See comment on Pintner and Keller article, No. 120. Ambiguity also lies in the fact that M.A. comparisons are based on children by grades rather than on children of known ages.

119. Pintner, R. "Results obtained with the non-language group test." *Jour. Educ. Psych.*, 15: 1924, 473-483.

Dividing 1361 twelve-year-olds into those with Anglo-Saxon, Italian, German or Dutch, and miscellaneous names, it was found that the only two groups showing a significant difference on the Pintner Non-Language Test were the Anglo-Saxon and Italian. The author concludes that it is significant that differences obtain between Italians and Anglo-Saxons on a non-verbal test, even though they are not so large as differences on a verbal test. The difference was about .25 of the S.D. of Anglo-Saxons.

Comment: The result is significant, but helps little in the interpretation of the far larger differences found on verbal tests.

120. Pintner, R., and Keller, R. "Intelligence tests of foreign children." *Jour. Educ. Psych.*, 13: 1922, 214-222.

Second-grade children tested upon the Youngstown Children's Bureau Binet Revision and the Pintner Non-Language Test were divided into 49 of English-speaking parentage and 56 of foreign-speaking parentage.

Seventy-three percent of the English-speaking, and 82 percent of the foreign-speaking were higher on the Pintner than on the Binet.

Using the median mental age on at least three performance tests for each child, 52 percent of 45 English-speaking children, and 75 percent of 95 foreign-speaking children were better on the performance tests than on the Binet.

The average I.Q. of 367 English-speaking children was 92, and of 674 children from foreign homes, 84.

The authors conclude: "Children who hear a foreign language at home test lower as a rule when given the revisions of the Binet Test than when given tests which require a minimum knowledge of English. . . . Those children who hear a foreign language in their homes may suffer a serious handicap when tested only by the revisions of the Binet Test." (See also data from Pintner and Keller in the tabled summary by Goodenough, No. 97.)

Comment: The fact that the non-language and the performance tests are measuring different abilities from those sampled by the Binet Test makes the interpretation ambiguous.

121. Saer, D. J. "An inquiry into the effect of bilingualism upon the intelligence of young children." *Jour. Exper. Ped.*, 6: 1922, 232-240.

The phase of the experiment of especial interest to our problem concerned boys of 7 to 10 at Aberystwyth, (selected from 260 boys of 7 to 14). The exact number is not stated for this portion of the investigation.

The Stanford Binet, plus non-overlapping questions in Binet's 1911 scale and in Burt's version, was given the subjects. The tests were carefully translated into Welsh, and two years later given to boys in the group who were bilingual, *i.e.*, boys who spoke English fluently, but whose "mother-tongue" was Welsh. To those who were monoglot, *i.e.*, whose mother-tongue was English, the same English tests were repeated in two years. The former group showed a gain in M.A. two and two-thirds months greater than that of the second group.

Comment: This section seems inadequately reported. Numbers are not given, nor the method of computing I.Q. However, the excess gain seems significant, since elsewhere the initial I.Q.'s of the two groups are seen to have the same medians—100 and 99.

122. Sunne, D. "Comparison of white and negro children in verbal and non-verbal tests." *School and Soc.*, 19: 1924, 469-472.

Negro children did slightly less poorly upon the Myers Mental Test than upon the National. The author concludes that "it is difficult to determine how much racial differences and how much differences in school training contribute to the divergence in test results."

123. Walters, F. C. "Language handicap and the Stanford Revision of the Binet-Simon tests." *Jour. Educ. Psych.*, 15: 1924, 276-284.

One hundred and sixty-five children in a New York school of a foreign district with an average I.Q. of 96 and average C. A. of 12-8 were divided into those hearing chiefly foreign language and those hearing chiefly English at home. The Stanford-Binet items on age levels

from 10 to 16 were divided into language and non-language tests, and I.Q.'s computed for the children with and without inclusion of the "language" tests.

Comparing I.Q.'s by the two methods, the author concludes that there is a language handicap of from 6 to 8 months of mental age for children of 13 coming from foreign-language-speaking homes.

Comment: The most valid and most reliable of the Binet test elements are the verbal ones. Hence a convergence of the groups on the "non-language" tests might easily occur, even if there were no language handicap among the children. Thus, the data of this study provide nothing crucial.

4. Intelligence and Schooling

Simple correlations between amount of schooling and intelligence of course contribute nothing crucial, since such relationships might or might not be due to selection. It is only when differences in intelligence appear among comparable individuals for whom the amount of schooling is the only primary difference that definite conclusions can be drawn.

124. Bagley, Wm. C. *Determinism in Education.* Baltimore, 1925, 194 pp.

Chapter IV. Do good schools pay? Inter-correlations are reported between the following data for the states:

1. Sandiford's 1923 distribution of states by percent of A and B men in Army Alpha.
2. 1920 census report on literacy.
3. Reeder's 1923 rankings of percent capita circulation of 10 magazines.
4. Proportion of population in *Who's Who* 1850-1880.
5. Knauth's distribution of per capita income, 1922.

Using the 26 states having over 55 percent native population, these r's range from .73 to .91.

The correlations between Army Alpha test rank and Ayres School Index (the latter computed for various decades) are also reported as: 1880, .83; 1890, .82; 1900, .89; 1910, .87; 1918, .83.

The author points out that Alpha ranking correlates highest with school ranking in the decades that the r's should be highest (*i.e.,* when the army draftees were in school). "Recent school conditions," he says, "have not affected the present dominant generations, but school conditions 20, 30, and 40 years ago did have a profound effect if our figures tell the truth."

It is also shown that the average correlation of four of the above measures with one another, "when the influence of school opportunity

is cancelled," is only —.09. The author ascribes this decrease to the "predominantly causal influence" of school opportunity.

Comment: It must be kept in mind that the high inter-correlations reported are based upon group *averages* and hence do not represent the true relation between the variables in question. Also, the average of the partial correlations with School Index constant is a highly ambiguous measure, because it is not known whether a cause or an effect is being partialled out. It seems fair to say that no data are reported here which could not as readily be interpreted in favor of heredity as environment.

125. Burt, C. *Mental and Scholastic Tests.* London, 1921, 432 pp. (Sec. 6, pp. 175-184. "The relations between mental ability and educational attainments.")

(1) For 689 London school children of ages 7 to 14 a correlation coefficient was computed to predict mental age on Burt's revision of the Binet scale from a composite of educational age, chronological age, and mental age on a "reasoning test" devised by Burt. The equation was: $B = .54S + .33I + .11A$, where S was equal to educational age (or schooling), I was equal to standing on the reasoning test, and A was equal to chronological age. The author attributes to every factor a "causal" influence equal to its proportional weight in the regression equation.

(2) In another section, the author cites 10 cases out of the total group showing an E.Q. 30 percent or more below the I.Q., but no child has an I.Q. so far below his E.Q.

Comment: (1) See critique of this study by Holzinger and Freeman summarized farther on (No. 128).

(2) The data above indicate that I.Q. places an upper limit upon the E.Q., and illustrate a method by which test data can be made to yield more than mere group treatment reveals.

126. Ellis, R. S. "A comparison of the scores of college freshmen and seniors on psychological tests." *School and Soc.*, 23: 1926, 310-312.

One hundred twenty-one Syracuse students who took the Miller Mental Ability Test when they were freshmen and three years later when they were seniors showed no improvement in average score. It was found, however, that gains of approximately .5 S.D. of the freshmen distribution were made during three years on the following tests: (1) sentence completion, (2) artificial language, and (3) opposites.

The author attributes these gains to "transfer of training" during the college course, rather than to any growth of fundamental capacities.

127. *Gordon, Hugh. *Mental and Scholastic Tests Among Retarded Children.* London Board of Education, Ed. Pamphlets, No. 44, 1923, 92 pp.

This contribution consists of four separate studies.

Study I. "Physically Defective Children"

In two schools for the physically defective, 184 children whose records of school attendance could be traced had been in attendance only 48 percent of the time since the age of five. (The corresponding percent for children in ordinary elementary schools is 88). The average I.Q. of the children (who were nearly all between 6 and 12 years old) was about 85. I.Q. and E.Q. (based upon the Ballard one-minute tests in speed of reading, adding, and subtracting) both correlated with percent of school attendance .31 ± .07, which fact is taken as evidence that school attendance affects both I.Q. and E.Q. The negative rank-order correlations of I.Q. and E.Q. with C. A. were also taken as evidence for the influence of schooling. These were:

I.Q. and C. A.

73 pupils in one school, — .37 ± .07
83 pupils in one school, — .42 ± .06

E.Q. and C. A.

83 pupils in one school, — .29 ± .07

Comment: May not the low correlations of attendance with I.Q. and E.Q. signify that achievement is far more determined by intelligence than by attendance? This possibility is strengthened by the fact that an r of .78 between I.Q. and E.Q. is only reduced to .76 when attendance is partialled out.

Nothing in the data, of course, can show whether the small r actually obtained between I.Q. and attendance is due to the influence of attendance on I.Q. or the influence of I.Q. on attendance.

The possibility is not dealt with that the slight inverse relationship between I.Q. and C. A. and between E.Q. and C. A. might be due to some type of selection.

Study II. "Canal-Boat Children"

This study deals with 76 children of ages 5 to 13 (except two of 4 years and one of 14) living on English canal boats. On the basis of a Ministry of Health report, the health and nutritional status of these children seems average. They differ from other children in type of home environment and in the amount of their schooling, which is only about 4 or 5 percent of normal.

The Stanford Binet Test and Ballard Achievement Test were administered. The rank-order correlation between C. A. and I.Q. was — .76 ± .03. Taking 22 families in which from two to four sibs were tested, the inferiority of the older siblings was also very apparent.

Twenty-two oldest sibs averaged close to 60 I.Q.; 22 second sibs close to 72; 11 third sibs close to 77; and 5 youngest sibs about 89.

Comment: The C. A.'s of these five youngest varied from 4-7 to 7-3. The median age of the oldest sibs was 12. A drop of a few points may be accounted for by a faulty scale, but evidence seems to point to an influence of either schooling or home environment. Selection was not considered, but there seems no special reason to suspect it. It is significant that I.Q. and E.Q. correlate nearly as highly (.715) in this group as in the former group (.785), although attendance was only 4 to 5 percent. This certainly suggests that achievement is more dependent upon intelligence than upon attendance.

Study III. "Gypsy Children in Surrey and Kent"

Gypsy children of ages 5 to 13, inclusive, who had attended school an average of 35 percent of the time since the age of five, were tested by the Stanford Binet and Ballard achievement tests.

I.Q. and E.Q. correlated $.78 \pm .03$ on 60 cases, and this r still remained at .78 when percentage of attendance was partialled out.

I.Q. and percent attendance since age of five correlated $.37 \pm .06$ on 82 cases, and E.Q. and attendance correlated .29 on 60 cases.

C. A. correlated with I.Q. $-.43 \pm .06$ on 82 cases, and with E.Q. $-.37$ on 60 cases.

Inverse relationships, as in the case of the canal-boat children, were also apparent between I.Q. and C. A. of siblings, though not to so great an extent. The average of 7 third children was about 81, of the 21 second children about 78, and of 21 oldest children about 72. The third children averaged about 7 years C. A. and the oldest about 11.

Comment: The same comment applies here as in the study on canal-boat children.

The author appears to claim too much when he writes: "It appears that for all practical school purposes the three simple scholastic tests may be substituted for the mental tests in the case of such children as have been tested." This does not follow from a correlation no higher than .78 between I.Q. and E.Q.

Study IV. "Backward Children"

Girls of ages 7 to 14 in two London classes for backward children had an average I.Q. of 70 to 75 and had attended school an average of 67 percent of the time. For 76 of them, attendance correlated $-.08 \pm .08$ with their I.Q.'s and .15 with their E.Q.'s. C. A. correlated $-.47$ with both the I.Q.'s and the E.Q.'s of 67 of them. Since their average attendance, despite their low I.Q.'s, was higher than that of the subjects of the previous sections, the conclusion is drawn that this group has low *native* intelligence.

The chief correlations of the study are summarized below.

	I.Q. and C.A.	E.Q. and C.A.	I.Q. and Attend.	E.Q. and Attend.	E.Q. and I.Q.
Physical defectives	−.422	−.288	.313	.313	.785 and .707
Canal boat........	−.755	....	....	....	.715
Gypsy..........	−.566	−.374	.283	.289	.784
Backward........	−.477	−.466	−.162	.273	.784

The author concludes: "It is quite evident that, although the mental tests used do undoubtedly test some kind of ability or abilities, such abilities are not developed without schooling or its equivalent, and as a consequence the tests do not evaluate them apart from schooling, except *perhaps* in the case of children under six or seven years of age."

Comment: The following types of evidence were presented:

(1) High r's between I.Q. and E.Q.

(2) Negative r's between I.Q. and C. A.

(3) Slight positive r's between I.Q. and attendance.

The first type is more likely to mean that E.Q. depends upon I.Q. than vice versa, especially since the correlations remain just about as high when percent of attendance is partialled out.

The third type shows low r's between I.Q. and attendance, but upon either a nature or a nurture hypothesis this would perhaps be expected, if the variability of attendance were not very great. (S. D. of attendance was mentioned only in the case of the "backward class" children, where it was 15 percent.) Thus, differences in attendance do not account much for the differences among these children, but the very low percent may or may not have resulted in a large cumulative effect upon the I.Q., or the whole r may be due to selection.

The second type of evidence suggests strongly that environment has had an effect, although it is impossible to assume, as the author seems to do, that the whole effect can be attributed to paucity of schooling as isolated from home environment. The author has not considered the possibility of selection in the upper ages, but the fact that in most of his groups there are about as many children in the upper ages as in the lower makes selection seem unlikely. The fact that in canal-boat families a drop in I.Q. occurs from about 90 to 60 in children of the same families from ages 6 to 12, and that a similar drop of from over 80 to 70 occurs in gypsy families, reënforces Gordon's theory, as do the lower negative r's, in classes of physical defectives and backward children, between I.Q. and C. A. Thus, the very extreme environment of the canal-boat children would seem to have a cumulative effect of about 5 points a year over at least six years.

128. Holzinger, K. J., and Freeman, F. N. "The interpretation of Burt's regression equation." *Jour. Educ. Psych.*, 16: 1925, 577-582.

The authors criticise Burt's famous regression equation purporting to show that Binet test scores are conditioned to the extent of more than one-half by schooling. Burt was in error in assuming *a priori* that his reasoning test was a good measure of native intelligence, that a test giving a good prediction of school achievement was necessarily a poor measure of intelligence, and that the multiple regression technique provided a flaw-proof mathematical statement of the causal relationships of a group of variables to another variable. This last criticism is driven home by a regression equation computed by the authors upon Burt's own data by which the chronological age of London school children is estimated by combining .15 Binet-Simon mental age; .51 schooling; and .03 mental age on the reasoning test. The ironical suggestion is made that to keep children chronologically young, we should deprive them of schooling and arrest their mental development.

See also the summaries in this section of Burt's original contribution (No. 125), of Thomson's answer (No. 130) to the present critique, and of the rejoinder by Holzinger and Freeman (No. 129).

129. Holzinger, K. J., and Freeman, F. N. "Rejoinder on Burt's regression equation." *Jour. Educ. Psych.*, 17: 1926, 384-386.

"If the capacity which is measured in the Burt score is intelligence, Thomson's interpretation [of the Burt equation] would seem to imply not only that degree of attainment in school does not affect intelligence, but also that intelligence does not affect the degree of attainment in school. Intelligence in this view must be some sort of useless ornament, enabling us to work certain puzzles which the psychologists have devised, but of no conceivable value in a practical work."

130. Thomson, G. H. "The interpretation of Burt's regression equation." *Jour. Educ. Psych.*, 17: 1926, 301-308.

The author attempts to defend Burt (and his own previous published acceptance of Burt's conclusions) against the attack of Holzinger and Freeman. The sum total of his argument is that while cause and effect cannot be argued from a correlation *per se*, the most reasonable interpretation is that Binet score is affected by school work but that Burt's test is not.

(See summaries Nos. 125, 128, and 129.)

Comment: This defense leaves the general field of the effect of schooling upon mental age as unsettled as it was before Burt published his widely quoted equation. Data are needed which are more crucial than any merely giving correlations between M. A. and E. A.

131. *Thorndike, E. L. "Mental discipline in high-school studies." *Jour. Educ. Psych.*, 15: 1924, 1-22, 83-98.

Using as subjects 8564 pupils in Grades IX, X, and XI, and allowing for practice effect and superior gains of the ablest on the Inst. of Educ. Research Tests of Selective and Relational Thinking, Thorndike reports the following results: 11 points for the year are gained by the average white pupil; 1.5 points for the year by the average colored pupil; 22.5 points, or 11.5 points above average, by virtue of taking three courses in science and one in mathematics; 19 points by virtue of taking one course each in arithmetic, bookkeeping, stenography, and typewriting; 10.5 points for one course each in cooking, sewing, dramatic art, and physical education.

By the best one percent in initial ability 20.5 points are gained; by the lowest one percent in initial ability, 1.5 points.

Also, with respect to individual subjects, if a pupil takes one of these: algebra, geometry, trigonometry, Latin or French, instead of one of these: cooking, sewing or stenography, the best estimate of his increment of gain is a little under 2.5 points, far less than the difference in gain of average negroes and whites, or of superior and inferior whites.

The author points out what he considers a lack of justice in refusing to permit pupils to offer subjects like stenography for college-entrance units. "The intellectual values of studies should be determined largely by the special information, habits, interests, and ideals which they demonstrably produce."

132. Wechsler, D. "On the influence of education on intelligence as measured by the Binet-Simon tests." *Jour. Educ. Psych.*, 17:1926, 248-257.

The coefficients of variability in Burt's and Terman's original Binet material at ages 5 to 14 of mental age are studied. The conclusion is drawn that education makes children "more alike" after age 10 or 11. Apparently, the possibilities of ambiguity due to selection and test standardization are not considered.

133. Willard, D. W. "Native and acquired mental ability as measured by the Terman Group Test of Mental Ability." *School and Soc.*, 16:1922, 750-756.

On the basis of this test, applied to 216 high-school pupils twice, (seven and one-third months apart), the author concludes that school training contributes to mental growth an amount equal to "a little over one-half the growth registered by the tests." If such a rate were kept up, by the end of the high-school course the children could be "classed as a group of near geniuses."

Comment: The study entirely overlooks probable practice effect, which on group tests is usually nearly equal to the gains here attributed to school training.

134. *Woolley, Helen T. "The validity of standards of mental measurement in young childhood." *School and Soc.*, 21:1925, 476-482.

The subjects were 43 children in the Merrill-Palmer School between 2-6 and 4-6 at first test, and 33 children, from comparable homes on the Merrill-Palmer waiting list, between 2-6 and 5-2 at first test.

The Stanford Binet was administered to all children (presumably before entering the school, in the case of the Merrill-Palmer pupils). Retests were given from 7 to 14 months later (median interval about 9 months for each group). Pupils attend the school from 9 A. M. until the middle of the afternoon, and are controlled by a definite educational program.

Sixty-three percent of the Merrill-Palmer group increased more than five I.Q. points, while 18.5 percent decreased more than five I.Q. points. Corresponding figures for the waiting list were 33 percent and 36 percent. Thirty-three percent of the Merrill-Palmer pupils increased 20 points or more; 10 percent decreased 20 points or more. Corresponding figures for the waiting list were 6 percent and 16 percent. Actual I.Q. distributions are not given. The average I.Q. of 19 cases tested a third time was "a little lower" than on the second test, but 18 pupils did not show an average decrease after leaving the school.

Comment: From Woolley's figures the writer finds an average increase of 13.6 points for the Merrill-Palmer pupils and an average decrease of 1.6 points for the waiting list.

135. *Woolley, Helen T., and Ferris, Elizabeth. *Diagnosis and Treatment of Young School Failures.* Washington, 1923, 115 pp. (Bur. of Educ. Bull., No. 1.)

This study deals with 16 children, C. A. 6-7 to 10-4, and I.Q. 75 to 95 at the beginning of the experiment. Not one had covered the prescribed first-grade work. The group constituted the first "observation class" of Cincinnati.

Exceptionally good training was given these children (apparently for two years) in a special class. They were tested from year to year by the Stanford Binet, performance tests, and educational tests.

"In every instance, except those of the children who were ultimately shown to be feeble-minded, and one of those who was suffering from special disabilities, the intelligence quotient rose while the child was in the observation class and fell after he left it."

Ten cases showed increases from 3 to 18 points, with an average of 8. "The reëxaminations of these same children after a year to a year and a half, in a regular grade, show an equally uniform fall in intelligence quotient." Decreases varied from 3 to 19, with an average of 8.

Comment: Environmental influence is clearly suggested by the data; but the cases are too few in number to establish such influence with certainty.

136. Yerkes, R. M. (editor) *Psychological Examining in the United States Army. National Academy of Science, Memoirs,* 15: 1921, 890 pp. (Ch. X. "Statistics on Schooling.")

Median schooling is seen to correspond to intelligence medians of the following groups (except northern negroes were better schooled, but less intelligent than the foreign draft).

	Schooling	Median Alpha
Officers	14.7	139.2
White draft (native-born)	6.9	58.9
White draft (foreign-born)	4.7	46.7
Negro draft, northern	4.9	38.6
Negro draft, southern	2.6	12.4

For members of these five groups who had had equal schooling, differences still persisted on Alpha and Beta. For example, among those who had from five to eight years of schooling:

	Median Alpha	Median Beta
Officers	107.0	....
White draft, (native-born)	51.1	53.3
White draft, (foreign-born)	47.2	52.7
Negro draft, northern	37.2	41.5
Negro draft, southern	16.3	35.1

Six hundred sixty officers with schooling of eighth grade or less had a median Alpha score of 107.3, and 14,000 native-born white draftees had a median of 97.4, which shows "that education is certainly not the chief conditioning factor."

On the basis of r's from a number of groups: "In an unselected group (*i.e.,* including those men who would ordinarily be considered too illiterate to take Alpha) the correlation with the number of years of their schooling approximates .75; in an Alpha group (*i.e.,* excluding illiterates) the correlation coefficient approximates .65." The range in schooling for these r's is from none to eight years of college (white native-born); and from none to four years of college in a group of native-born and foreign. The r between Beta and schooling for the 653 men is .67.

The data are interpreted as meaning that the "theory that native intelligence is one of the most important conditioning factors in continuance in school is certainly borne out by this accumulation of data."

Comment: The data seem to provide evidence of *some* influence of nature; but most of the data could be explained either by nature or by nurture.

5. Relation to Physique, Health, or Physical Environment

Studies of this kind, of course, furnish significant evidence in favor of acquired changes in mental level only when mental changes are demonstrated in connection with physical changes or when mental differences are found associated with physical ones in comparable experimental and control groups. A mere correlation between disease or defect and intelligence cannot establish a causal relationship, since selection or common causes might account for such a correlation. However an *absence* of correlation between mental level and physical condition is significant, since such a result dispenses with the necessity for considering causal relationships at all.

137. *Dawson, S. and Conn, J. C. M. "Effect of encephalitis lethargica on the intelligence of children." *Arch. Dis. Childhood,* 1:1926, 257-368.

The following results were based upon Binet tests (Burt's revision) given to 46 children in non-acute stages of *encephalitis lethargica* at the Royal Hospital in Glasgow.

(1) The patients in question had an average I.Q. of 84.6, as contrasted with 90.5 for 974 other children tested in the hospital. The difference is 5.9 ± 1.3.

(2) Twenty-three patients tested less than 12 months after the onset of the disease and 23 tested from 12 months to 5 years 4 months after onset had average I.Q.'s of 89.5 and 79.8, respectively. The difference is 9.7 ± 2.3.

(3) Thirty children retested after 7 to 36 months had an average I.Q. of 87.7 on the first test and 76.1 on the second test. The difference is 11.6 ± 2.1. In 10 cases there was actual retrogression of M. A. and in 20 cases some gain, but in 12 of these the gain was negligible. In 27 cases there was *less gain* than normal for children of the given I.Q.'s during the interval between tests, and in only five cases was the ratio of gain to expected gain as high as 75 percent. The correlation between rise above expected gain and interval between the two tests is $-.55 \pm .08$. Thus, mental development is seen to be arrested by the disease, although half of 8 patients who were given three tests showed some improvement on the third test.

(4) A group of 23 subjects with available siblings averaged 85.6 on their first test, and 20 of these averaged 76.0 on a retest. Their 27 tested siblings averaged 96.0. Thus, the patients differed from their siblings 10.4 ± 2.2 on the first test and 19.6 on the retest.

(5) Sixteen patients with the Parkinsonian syndrome and 30 without it averaged 84.9 and 84.5, respectively.

(6) Twenty-nine patients with emotional disturbances and 17 without them averaged 84.2 and 84.8, respectively.

138. *Fernald, M. R., and Arlitt, A. H. "A psychological study of a group of crippled children of various types." *School and Soc.*, 21: 1925, 449-452.

In the Cincinnati School for Crippled Children, 49 pupils tested by the Stanford Binet had an average I.Q. of 83.9. Eighty-nine siblings of these pupils had an average I.Q. of 89.2. The difference, approximately five points, is possibly due to the restricted early activities of the crippled children, although no relationship between age of onset of malady and I.Q. was found among 194 cases, if spastic birth paralysis cases were not considered. Twenty-seven of these latter had an average I.Q. of only 69. A correlation of $.52 \pm .07$ was found between the scores of 49 cripples and those of their normal siblings.

Comment: The correlation immediately above is ambiguous and spuriously high because the I.Q.'s of all normal sibs of each family were averaged to pair against the I.Q.'s of the crippled pupils.

139. Heron, David. *The Influence of Defective Physique and Unfavorable Home Environment on the Intelligence of School Children.* Cambridge University Press, 1910.

(As summarized by Heron in "Questions of the Day and of the Fray, No. II.")

"Particulars were given for 4286 boys and 4474 girls regarding mental capacity, age, standard [grade], height, weight, and condition of teeth, and in certain schools only regarding state of nutrition, condition of the clothing, state of cleanliness, power of hearing, and condition of the cervical glands, tonsils, and adenoids."

"The conclusion was reached that home environment . . . . could not be the chief determining cause of the differentiation of intelligence, nor was defective physique its source."

The possibility was also pointed out that the small relationships found might be an "indirect effect of race and stock."

140. *Laslett, H. R. *Experiments on the Effects of the Loss of Sleep.* (Ph.D. dissertation, Stanford University, 1926, 148 pp.)

The subjects were four college men who lived on a "40 percent sleep ration" for five days and five college men who went entirely without sleep for 72 continuous hours. A code writing test, an addition test, and Part I of the Thorndike College Entrance Test were used to measure changes in mental efficiency.

The author concludes: "Long periods without sleep cause temporary loss in the ability of the individual to perform tasks. The more the task involves the higher mental processes, the greater is the loss."

During the 40 percent "ration" period "the losses were not as great as those in the period during which other subjects obtained no sleep."

Comment: The general conclusion seems sound, but the results are difficult to evaluate precisely because (1) the subjects are so few, and (2) the losses in efficiency are measured in terms of *percent* of *original* efficiency and hence are not comparable on all the tests.

141. *Lowe, Gladys M. "Mental changes after removing tonsils and adenoids." *Psych. Clinic,* 15: 1923, 42-101.

The subjects were 60 children with defective adenoids and tonsils, of whom 35 were operated on after the first test.

	Non-Operated Cases	Operated Cases
First I.Q.	83.0	89.2
I.Q. a year later................	85.4	90.7

No effects of diseased adenoids or tonsils upon the I.Q. were manifest.

142. *Olesen, R., and Fernald, M. R. "Endemic goiter and intelligence." *Public Health Reports,* 41: 1926, 971-986.

Measurements of 3800 sixth-grade Cincinnati children, both white and colored, showed equal Otis Test levels for those with thyroid enlargement and for those with normal thyroid glands in the white group, and Otis Test levels six to eight points higher for colored children with thyroids enlarged than for colored children with thyroids normal. The discrepancy in the colored group may be partly due to small numbers and consequent errors of sampling.

When intelligence was considered in the white group with respect to size of thyroid enlargement, the results were these:

	Boys		Girls	
	Percentile	N	Percentile	N
Slight	46	387	51	484
Moderate	50	42	50	105
Marked	32	16	45	53

This suggests a lowering in the case of marked enlargements, but the cases are too few in number to be certain.

The conclusion is: "There is no evidence that either boys or girls with slight or moderate degrees of enlargement differ significantly in intelligence from those with normal thyroids." While marked enlargement probably had some effect, there is "no justification for an assertion of general relationship between thyroid enlargement and lower intelligence."

143. *Paulsen, Alice. *The Influence of Treatment for Intestinal Toxemia on Mental and Motor Efficiency.* New York, 1924, 45 pp. (Archives of Psych., 11: No. 69.)

Thirty experimental subjects and 30 control subjects, nearly all of them young women in their twenties, were matched fairly closely for Terman Group Test scores. The experimental group were suffering from intestinal toxemia at the beginning of the experiment, but improved in this condition during the month of the experiment, during which remedial treatment was given.

A battery of 17 tests was administered to both groups at the beginning and end of the experiment, and the net average gain of the experimental group was computed. This tended to be from about one-sixth to one-third of the S. D. of the experimental group on the 12 *mental* tests, and from about a half to a whole S. D. on five *motor* tests. The average net gain was 8.8 times its P.E. on the whole battery.

The mental tests used ranged in complexity from opposites and hard directions down to cancellation and color naming.

144. *Pearson, K. "On the relationship of intelligence to size and shape of head." *Biometrika*, 5: 1906-07, 105-146.

A group of 1000 Cambridge men were rated on a four-point scale according to honors or pass degree. Over 5000 school children were rated on a six-point scale for intelligence. Boys and girls were classified separately. Measurements of the school children were reduced to a standard age of 12 by aid of the average growth curves. Correlation ratios were computed between these ratings and certain physical measurements as follows:

	Cambridge		Boys		Girls	
	Cor. Ratio	P.E.	Cor. Ratio	P.E.	Cor. Ratio	P.E.
Intelligence and cephalic index	−.061	.021	−.041	.014	.067	.014
Length of head	.111	.020	.139	.014	.084	.014
Breadth of head	.097	.021	.109	.014	.111	.014
Auricular height	......	....	.073	.014	.055	.014

145. *Rogers, M. C. *Adenoids and Diseased Tonsils; Their Effect on General Intelligence.* New York, 1922, 70 pp. (Archives of Psych., No. 50.)

Twenty-eight boys (ages 6 to 14) having diseased adenoids and tonsils were tested by the Stanford Binet Test the mornings upon which they had adenoid and tonsil operations, again six months later, and again 12 months later. Their I.Q.'s averaged 2.25 points higher after six months. But 28 control boys, comparable to the above groups for age and diseased condition of adenoids and tonsils, gained 3.29 points on a second test after six months, though they had had no operation. After

12 months, 21 boys still available in the operated group had risen 3 points and their 21 paired controls had risen 6 points. No significant correlation was found between age (presumably length of time defect had been present) and rise in I.Q.

Two other groups of children in a large public school showed no difference in average I.Q. when separated into those who had diseased adenoids and tonsils and those who had normal adenoids and tonsils.

Comment: The study seems to establish definitely that diseased adenoids and tonsil conditions have no unfavorable effect upon intelligence as such. The article would contain additional interest if the I.Q.'s could have been watched over a longer period of time.

146. Smillie, W. G., and Spencer, Cassie R. "Mental retardation in school children infested with hookworms." *Jour. Educ. Psych.*, 17:1926, 314-321.

The Otis Test and the National Intelligence Test were given to 118 children in a hookworm area in an Alabama county. Only Otis I.Q.'s were used in the analysis, but these were "verified" by one or two N.I.T. scores. The Stoll ova count method was used at the Andalusia hookworm research laboratory to estimate the number of hookworms harbored in each child. Since all children lived on neighboring farms and had white American parents, the authors believe that selection probably does not enter.

TABLE OF RESULTS

I.Q. Group	N	Estimated Number of Hookworms
105–115	4	30
95–104	18	47
85– 94	45	67.8
75– 84	36	172
65– 74	15	281

Intensity of Infection	N	Mean I.Q.	P.E.
Negative	17	90.2	1.53
Very light (1–25)	40	88.3	1.06
Light (26–100)	27	86.4	1.2
Moderate (101–500)	23	84.1	1.07
Heavy (501–2000)	10	76.3	1.74

It is stated that this study contributes data which previous hookworm studies have not provided concerning the effects of various intensities of hookworm infection.

Comment: The authors' belief that selection does not affect results is not demonstrated. It still remains possible that children of incompetent parents were more likely to be exposed to hookworm.

147. Terman, L. M. *The Intelligence of School Children.* Cambridge, 1919, 313 pp.

For 27 children retested after an operation for tonsils or adenoids, the average gain in I.Q. was two points. There were 10 losses and 17 gains, no gain larger than 14 points, and only two larger than 10 points. The results "suggest that adenoids and diseased tonsils may give a child an exaggerated appearance of dullness."

148. *Thorndike, E. L., McCall, W. A., Chapman, J. C. *Ventilation in Relation to Mental Work.* N. Y., 1916, 83 pp. (Teachers College, Columbia University, Contrib. to Educ., No. 78.)

Conditions of ventilation had no effect upon the mental efficiency of over 100 college students, measured on the following tests: naming 100 colors; cancelling 2's for 60 seconds; cancelling 3's for 60 seconds; 50 hard opposites; addition of 10 digits for 10 minutes; mental multiplication of 3-place numbers by 3-place numbers for 20 minutes; 20 minutes of typewriting, and other miscellaneous tasks.

The conditions ranged from 68° F., 50 percent relative humidity, and 45 cu. ft. of air per person per minute to 86° F., 80 percent relative humidity, recirculated air and stagnant. Favorable conditions and poor conditions were alternated daily in some experiments and conditions were held constant for five consecutive days in others.

149. *Thorndike, E. L., Ruger, G. J., McCall, W. A. "The effects of outside air and recirculated air upon the intellectual achievement and improvement of school pupils." *School and Soc.,* 3:1916, 679-684.

(See Thorndike and Ruger, No. 150.)

150. Thorndike, E. L., and Ruger, G. J. "The effects of outside air and recirculated air upon the intellectual achievement and improvement of school pupils (a second experiment)." *School and Soc.,* 4:1916, 260-264.

The results of article No. 149 are also summarized here.

On two successive years experiments were made upon two classes of about 40 pupils each with respect to the relative influence of outside air and recirculated air upon mental and school efficiency. In the groups supplied with recirculated air occasional outside air was introduced, but steam consumption in classroom heat was diminished 50 percent. During one-half a school year the groups supplied with *recirculated* air showed a very slight tendency toward larger gains than the "fresh air groups" on tests of the intelligence and achievement type, and on school marks.

Comment: The slight tendency toward larger gains in the "recirculated air groups" is possibly accidental or due to errors of sampling.

It is significant, in any event, that these groups show no *diminution* in efficiency.

151. *Yerkes, R. M. (Editor) *Psychological Examining in the United States Army. National Academy of Sciences Memoirs,* 15: 1921, 890 pp. (Ch. XIII. "Influence of Certain Physical Conditions on the Intelligence Score.")

(1) At Camp Custer 178 men who had had the complete typhoid inoculation about 24 hours before were compared with the entire increment of the same month, 7,167 men. The 178 had a mean Alpha score of 67.7 and S. D. of 33.5. The 7,167 had a mean Alpha score of 67.1, and S. D. of 35.6. This comparison "is deemed more satisfactory" than one with 239 men of the same unit (not inoculated) who had a mean score 74.2 and S. D. of 35.3 because the latter "group contained a good many superior men who were being held over....because of possible fitness for officers' training camps." Seventy men "who themselves judged that their showing on the scale had been affected by their recent inoculation" had a mean score of 67.4 and S. D. of 33.5; 108 who "felt this to have had no effect" had a mean of 67.9 and S. D. 38.8.

(2) Data are reported showing recruits infected with hookworm to be distinctly inferior, and venereals somewhat inferior, but it is pointed out that selection is a possible explanation. In fact, 317 white venereals, who became infected after the test, were distinctly below unselected men at the same camp (Lee).

6. Relation to Age or Maturity

Selection complicates most studies under this heading (when scores of different age groups are compared). A few studies, however, give retests of the *same children* at different ages, and provide crucial evidence on the duration of mental growth. In most studies thus far made, however, we are not justified in assuming that the actual form of mental growth is shown in more than a very general way, since the scaling of most tests is not such as to enable us to say that one unit at a given level on the scale represents the same increment of ability as does one unit at another level.

A point of special interest concerns the average adult mental level as compared with the mental level of young people in their teens and still at school. Growth curves for young people usually extend to a level higher than that representing the *average adult mental level,* but whether this fact is due to actual decline of adult capacities or merely to the greater remoteness of school training in the case of adults is difficult to ascertain.

152. Ballard, P. B. "The limit of the growth of intelligence." *Brit. Jour. Psych.*, 12: 1921, 125-141.

Lumping results from an absurdity test given to elementary-school, secondary-school, and continuation-school pupils and evening-school adults, it is found that growth practically ceases at age 15.

The author admits that he can not be sure of having unselected groups, but believes that he has demonstrated no growth on this test after 16 because in secondary schools, where selection might be expected to raise the scores of older pupils, the maximum level was reached at age 15.

153. Brooks, F. D. *Changes in Mental Traits with Age.* New York, 1921. (Teachers College, Columbia University, Contributions to Education, No. 116.)

When 171 children in Grades IV to IX and between ages 9 to 15 were tested twice, a year apart, on a battery of 16 intelligence and achievement tests, there was an increase at all the ages. There was a slight tendency for the increase to be less in magnitude at 14 to 15 than at 9 to 10, but the significance of this is clouded by possible selective factors.

154. *Cobb, Margaret V. "The limits set to educational achievement by limited intelligence." *Jour. Educ. Psych.*, 13: 1922, 449-464.

The increase in Alpha scores from freshmen to senior standing in the high school which are due respectively to (1) mental growth together "with the temporary effects of instruction" and (2) elimination and the "more permanent effects of instruction" are roughly estimated.

The superiority of senior scores over freshman scores in high schools of eight states tended to be about 30 points. But the superiority of army recruits who had been seniors in high school over those who had been only freshmen in the high school was only 17 points. Thus, (1), above, seems to account for about half the difference between freshmen and seniors. But (1) seems to account for no significant superiority of seniors over juniors by a similar comparison. This superiority is much less than in other grade comparisons.

155. Hart, H. "The slowing up of growth in mental test ability." *School and Soc.*, 20: 1924, 573-574.

A total of 579 Iowa children born in 1908 attending intermediate, high, and part-time schools in Davenport, Iowa, was studied. As the tests were given in December, 1923, all the children were between 15 and 16.

The Army Alpha test gave an r with C. A. of .10 ± .03, which corresponded to a gain of 5.16 (or almost surely between 0 and 10 points,

allowing for 3 P.E. variation). If the gain were only 10 points a year since year 9, the average Alpha score would be 30 at age 9, whereas it is only 4 by Alpha norms. These facts are taken as a demonstration that growth is much slowed up at 15.

Comment: The possible factor of unequal units at lower and higher levels obscures the results somewhat.

156. Hopkins, L. T. *The Intelligence of Continuation School Children.* Cambridge, 1924, 132 pp.

All the 14- and 15-year-old children in five Massachusetts communities were tested on the Dearborn Intelligence Scale. These included both continuation-school and regular-school pupils. In the regular-school group 60 percent of 14-year-olds and 69 percent of 15-year-olds overlapped the medians for the same ages of the regular and continuation pupils combined. In the continuation-school group 18 percent of 14-year-olds and 28 percent of 15-year-olds overlapped these medians. The difference between ages 14 and 15 for the two groups combined was about 4.5 points.

Comment: Since the variability of the group is not reported, the significance of the age gain is not entirely clear.

157. Symonds, P. M. "Second approximation to the curve of the distribution of intelligence of the population of the United States." *Jour. Educ. Psych.,* 14:1923, 65-82.

By taking the medians and first and third quartiles Army Alpha scores for occupational groups (as given by Fryer) and weighting these according to the frequencies of the occupations listed in Table 14, p. 53, vol. 4 of the 1910 U. S. census, a mean Alpha score of 50.66 is obtained, a median of 48.41, and an S. D. of 43.00. For the army, a weighted composite median of the white draft (native and foreign-born) and the colored draft was 51.

The author thinks the Stanford Binet was standardized on selected groups, because occupations testing at median on Alpha test have an I.Q. of only 80 or 82 on the Stanford Binet, which would correspond to an M. A. of 13.2.

158. Teagarden, Florence M. *A Study of the Upper Limits of the Development of Intelligence.* New York, 1924. (Teachers College, Columbia University, Contributions to Education, No. 156.)

The subjects were 408 children of 12 to 20 at "Mooseheart," a cottage-plan institution for dependent children of Moose lodge deceased members. Feeble-minded children are not accepted there. The average I.Q. was 93.5, but this may be a little low, since 16 was used as the 'critical divisor' for the older ones. The author believes that selection from age to age is not operative, except in those over 18, because

"school mortality" does not exist in the ordinary sense. Children are not admitted after 14, but are kept and educated through their teens—by academic training if they can take it, by vocational training otherwise.

The Stanford Binet, Army Alpha, Pressey Senior Classification Test A, and Stenquist Mechanical Aptitude Test I were administered.

Three ways of scaling tests to make them comparable with themselves and each other at all ranges were tried, and gave almost identical results; so one only was used which assumed a normal distribution for the total population of children. The author points out objections to this assumption, but thinks it will answer for practical purposes.

With slight negative accelerations, the tests all show increases beyond 16; Binet at least to 18, Alpha to 16 or 16.5, Pressey to 17 or 17.5, Stenquist to 18, and Alpha to 16.5 or 17.5.

Comment: The possibility of selection does not seem entirely ruled out, although the results might perhaps be accepted tentatively until more crucial ones can be obtained in the future.

159. *Thorndike, E. L. "On the improvement in intelligence scores from fourteen to eighteen." *Jour. Educ. Psych.*, 14:1923, 513-516.

From 2638 to 3136 pupils in each of Grades IX, X, XI at first testing, and in Grades X, XI, and XII at second testing one year later served as subjects.

Form A of composite Institute of Educational Research tests was given to about half and Form B to the other half of the pupils in 1922; then the forms were reversed in 1923. The two forms were equated for relative difficulty, and a practice effect averaging 11.9 was found by taking scores of pupils (not counted in the subjects above) who took the test for the first time when the others were taking it the second time, and by comparing the scores.

The median gains for the pupils in Grades IX, X, XI in 1922 (after practice effect was allowed for) were 10.5, 11.7, and 11.5. The average gain was 11.1.

This is one-third of the mean square deviation of all the pupils on their first trial, "or about one-half of what their mean square deviation would be if they were perfectly measured." The estimate is made (how not stated) that the variability of all the pupils is at least half that of unselected 14-year-olds, who have an estimated mean square deviation of 23 months Stanford Binet M. A. Thus, the gains "may be set as equivalent to at least 10 months of mental age around 14."

Discussing the apparently constant rate of gain from Grades IX to XII, the author says "any decrease in gain with age, if such there be, is offset by the selection of those more capable of gain."

160. *Thorndike, E. L. "On the improvement in intelligence scores from thirteen to nineteen." *Jour. Educ. Psych.*, 17:1926, 73-76.

Several thousand high-school pupils from 13 to 18 in two cities were tested on I.E.R. Test of Selective and Relational Thinking. Selection was allowed for by finding an *upper* and *lower* limit of "superiority" in gains in one year (Form A and Form B) of younger pupils over older ones. Practice effect was also allowed for.

The gains of pupils from age 13 to 14 at various grade-placements is not significantly greater than those of pupils from age 14 to 15, etc., up to 18 to 19.

161. Thurstone, L. L. "A method of scaling psychological and educational tests." *Jour. Educ. Psych.*, 16: 1925, 433-451.

Two fundamental formulas are derived for establishing on an absolute scale the mean ability of an age-group when that of an adjacent lower age-group is known. Using these formulas upon the scores of 3000 London children from ages 3 to 14 tested by Burt on his Binet Revision, it is found that mental growth shows little signs of slowing up at age 14. The author concludes from the form of curves that it might "reach a limit in the early twenties or perhaps even at the age of 20, but it can hardly be extended to reach a limit much sooner than that."

Comment: The conclusion may possibly be distorted by selective factors of school mortality in Burt's data.

162. Whipple, G. M. "Endowment, maturity, and training as factors in intelligence scores." *Sci. Monthly*, 18: 1924, 496-507.

"In one Michigan city it was possible to test every 14-year-old and every 15 year-old child. The average scores for 14 and 15 years were 139 and 142 respectively; *i.e.*, the performances in these two ages were practically indistinguishable."

Increments from 8 to 9, etc. up to 14 to 15 are: 15, 15, 17, 16, 15, 9, 3. (The National Intelligence Test was used.) Numbers or school status of pupils were not stated.

163. Woolley, Helen T., and Fischer, Charlotte R. *Mental and Physical Measurements of Working Children.* New York, 1914, 247 pp. (Psych. Monographs, 18: No. 77.)

The subjects were native-born white children applying for work certificates at age 14, who had not been out of school longer than three months.

Mental tests selected from Whipple's Manual were administered twice, apparently to several hundred children a year apart (within two months plus or minus). Physical tests were also given (which are not commented upon here). Cancelling a's test; opposites; rote memory for 7, 8, and 9 digits; sentence completion (13 sentences); card sorting; form-board; and puzzle box (content was changed for second

test) were the mental tests used. The authors conclude that "the greater part [of the gain] can be attributed to the added year of maturity."

Comment: Results are presented in tabular form, but the standardization of the tests and data on dispersion are not adequate for precise interpretation. It is not possible, either, to separate practice effect from actual gain.

164. *Woolley, Helen T. *An Experimental Study of Children.* New York, 1926, 762 pp.

The treatment of data gathered ten years previously and reported in the 1914 and 1915 references is extended. Consecutive tests were given both to the working children and school children from ages 14 to 18. The study started out with 760 school children and ended with 152; and with 753 working children and ended with 511. A fresh influx of school children was added at age 16. While possible practice effect on test scores had not been adequately allowed for, it is significant that in the total group of school and working children combined: "The easier the mental task, the earlier an approximately adult status is reached. Thus, in memory an adult status is attained for the seven-place series at 15, for the eight-place at 16, and for the nine-place at 17 years." "Substitution" shows negligible gains beyond 16 years. Growth in these functions continues as long for the working children as for the school children.

Comment: All children tested at any age were used in the comparisons of age gains. It might have been more valuable to make comparisons using *only* the identical children throughout. The possibility that practice effect may account for a small or even large part of the gains cannot be overlooked. But the fact that gains occur in working children as long as in school children would indicate that *schooling* is not responsible for the gains.

165. Woolley, Helen T. "A new scale of mental and physical measurements for adolescents, and some of its uses." *Jour. of Educ. Psych.*, 6: 1915, 521-550.

Seven hundred fifty 14-year-old children were tested, and 680 of the same children were retested at age 15 after they had dropped out of school and been at work a year.

A series of physical and mental tests was given twice with a year's interval.

The opposites test was also given, but a harder list at age 15 than at 14; so it is not comparable.

Comment: The study is significant in showing that, even with children who have stopped school, there is some gain from 14 to 15. Practice effect is not guarded against, however.

Test	Gain 14th-15th Year		Q of Boys
	Boys	Girls	
Cancellation index	+37	+44	105
Cancellation (percent accuracy)	11	14	20
Memory, 7-place digits	3	4	16
Memory, 8-place digits	6	4	26
Substitution (4th page), (other pages seem to check)	6	9	40
Sentence completion (number correct)	1.1	0.7	2.8

166. Yerkes, R. M. (Editor) *Psychological Examining in the United States Army. National Academy of Science Memoirs,* 15: 3, 1921, 890 pp.

Chapter 14. "Relation of Intelligence to Age."

A group of 15,385 white officers from the "principal sample" had a median Alpha score of 150 for 24 cases of age 20 or less. The scores remained at about 146 from age 21 through 26, and then dropped, slowly at first, more rapidly later, to 120 in the group of 51 to 60. Eight cases of age 60 and over had a median of 140. The author points out that the unknown selective influences make the results ambiguous. Data from other scattered officer groups corroborate the form of curve, and data from over 1000 enlisted men, half each on Alpha and Examination a, show no decline from 20 to 30, thus confirming the results on officers. But in a group of 600 negro recruits the medians of men 21-23 and 29-31 rose from 34.8 to 40.5 (Alpha).

The following r's were found between age and Examination a:

5000 Med. officers (cf. below)	—.192	
2000 Med. officers (21-30)	—.063	(P.E. .013)
3000 Med. officers (30-60)	—.192	(P.E. .011)
146 Med. officers	—.011	
308 Infantry officers	—.120	

Chapter 11. "Intelligence of the Draft."

The principal sample of scores of the white draft, when transmuted from Alpha and Beta examinations into terms of mental age, have a median of 13.08 years.

7. Effect of Volitional and Emotional Traits

This is a most important topic, upon which little experimental literature can be found.

167. Bronner, Augusta F. "Attitude as it affects performance of tests." *Psych. Rev.,* 23: 1916, 303-331.

Using Binet's 1911 revision and tests from Healy and Fernald's *Tests for Practical Mental Classification,* the author reports case studies

of children tested at the Juvenile Psychopathic Institute who improved two or three years or more in M. A. on a second testing a few days or weeks later. Clinical descriptions make it evident that blocking of performance on Binet as well as on Healy-Fernald tests was often due to the following factors:

(1) Deliberate deception

(2) Recalcitrancy

(3) Sportiveness

(4) Emotional disturbances, including general depression, anger or resentfulness, fear, sheepishness, shyness, embarrassment due to onlookers, homesickness, mental conflict

(5) General excitement, including exceptional desire to do well

(6) Lack of confidence

(7) A combination of causes

Comment: The small numbers and subjective character of the clinical observations prevent the study from yielding precise conclusions; but the results seem important and suggestive, nevertheless.

168. *Hurlock, Elizabeth B. "The effect of incentives upon the constancy of the I.Q." *Ped. Seminary*, 32: 1925, 422-434.

One hundred forty-one eighth-grade and 132 fifth-grade children in two New York public schools were divided evenly into a control group, "praised" group, and "reproved" group on the basis of the National Intelligence Test, so that the mean and S. D. on the test were almost identical, as were the C. A., I.Q., and proportion of each sex in each group.

The National Intelligence Test, Scale B, Forms 1 and 2 were administered a week apart. The three groups took it together the first time, but took it in single groups the second time. The second time the control were merely given standard directions; but the "praised" were told they had been singled out from the entire class; the "reproved" were told they were a disgrace to the class. Results:

	Aver. I.Q. First Test	S.D. First Test	Increase Second Test	Increase in Units of P.E. of Diff.
Control	100.23	18.11	.69	.40
Praised	100.30	18.02	6.74	3.62
Reproved	100.43	18.04	6.88	3.72

Advantages are found, therefore, for both the praised and the reproved groups.

169. Hurlock, Elizabeth B. *The Value of Praise and Reproof as Incentives for Children*. New York, 1924, 78 pp. (Archives of Psych., No. 71.)

(For abstract, see summary No. 231.)

170. Jewett, S. P., and Blanchard, Phyllis. "Influence of affective disturbances on responses to the Stanford Binet test." *Mental Hygiene*, 6:1922, 39-56.

The authors conclude that I.Q.'s on the Stanford Binet are modified by psychoses, drugs, or emotional disturbances. Fourteen case studies are given in which very radical shifts of measured I.Q. can reasonably be attributed to disturbances of personality.

171. Sullivan, E. T. *Mood in Relation to Performance.* New York, 1922, 69 pp. (Archives of Psych., No. 53.)

No relationship was found among a group of adults between an introspective measure of cheerfulness and their performance on a battery of short mental tests. The conclusion was: "Extremes of mood act as distractions, the absence of any demonstrable effect on performance being due to the tendency of the individual to overcome resistance by increasing the output of energy."

8. Constancy of Intelligence

Studies upon stability of intelligence over short or long periods may show that whatever is measured is a permanent attribute, but can not show what determines it unless retests are analyzed in connection with *known changes* in environment. Nevertheless, it is important merely to know whether or not intelligence is stable, since a tendency toward *unstable* intelligence, if found, would be favorable to an environmental theory, even though *stable* intelligence would not be crucial evidence for a theory of heredity.

172. Baldwin, B. T., and Stecker, Lorle I. *The Mental Growth Curve of Normal and Superior Children Studied by Means of Consecutive Intelligence Examinations.* (University of Iowa Studies in Child Welfare, 1922, No. 1, 61 pp.)

For a group of 36 children who were given five successive Stanford Binet tests at average intervals of 13 months, 15 months, 10 months, and 5 months, the correlations between tests ranged between .74 (first and third test), and .92 (fourth and fifth test). The correlation for the first and fifth test was .82 ± .04. Correlations for somewhat smaller groups tested on two to four occasions ranged from .72 to .93.

173. Baldwin, B. T. "Additional data from consecutive Stanford Binet Tests." *Jour. Educ. Psych.*, 13:1922, 556-560; also in *Jour. Educ. Research*, 8:1923, 375.

(See Dickson's tabled summary, No. 177.)

174. Broom, Eustace. "Constancy of the I.Q." *School and Soc.*, 25: 1927, 295-296.

The Terman Group Test was given twice to 50 high-school pupils. The intervals between tests varied from 6 months to 35 months. On the first test the mean I.Q. was 110 and S. D. 12. On the second test the mean I.Q. was 110 and S. D. 10. Age 16 apparently is the basic C. A. used for pupils over 16. The r between test and retest was .86 ± .03.

175. Cox, Catharine M. *Genetic Studies of Genius, Vol. II. The Early Mental Traits of Three Hundred Geniuses.* Stanford University, 1926, 842 pp.

This represents one of the most significant studies of I.Q. constancy ever made. In the author's words: "The purpose of this study is to characterize a group of young geniuses with respect to certain mental traits. The subjects described are 301 of the most eminent men and women of history. The data discussed are the historical records of their heredity, their childhood, and their youth."

From the first 510 on Cattell's list of the 1000 most eminent individuals of history, 282 were selected who were born between 1450 and 1850. Hereditary aristocracy and nobility were excluded. These 282 formed the main group of the study from which the chief conclusions were drawn.

Upon the evidence of data collated from 3000 biographical sources, Professor Terman, Professor Maud A. Merrill, and the author estimated the I.Q.'s of the subjects, (1) in childhood up to age 17, and (2) in the period from 17 to 25. "The ratings were made on the basis of the standards and norms for intelligent behavior established by mental tests."

	Number	Aver. I.Q. up to 17 Yrs.	Same, Corrected for Regression	Aver. I.Q. 17–25	Same, Corrected for Regression
Artists and Writers (poets,	13	122	140	135	160
novelists, dramatists)	52	141	160	149	165
Writers (essayists, historians, critics, scholars)	43	139	160	148	170
Musicians	11	130	145	140	160
Soldiers	27	115	125	125	140
Scientists	39	135	155	152	175
Philosophers	22	147	170	156	180
Revolutionary statesmen	9	140	160	144	165
Religious leaders	23	132	150	145	170
Statesmen	43	135	155	142	165
Highest I.Q.'s	10	175	185	185	195
Lowest I.Q.'s	10	113	140	123	150
Total Group	282	135	155	145	165

Ten sub-groups dividing the subjects upon the basis of field of activity were formed. The average estimated I.Q. of members of these sub-groups (by all three raters) for both (1) the early period and (2) the later period of youth, are tabled below. Average I.Q.'s representing the probable level of the subjects after *regression of ratings toward the mean of the generality due to the unreliability of the data* had been allowed for are also tabled.

Data are also given upon character traits of the subjects in childhood, but these are not summarized here. The most significant conclusion of the author is that "the extraordinary genius who achieves the highest eminence is also the gifted individual whom intelligence tests may discover in childhood. The converse of this proposition is yet to be proved."

Comment: The validity of the correction applied for regression toward the mean of ratings based upon unreliable data is open to question, since it cannot be told, *a priori,* whether certain data were scanty (*i.e.,* "unreliable") because there actually *were* few indications of precocity in the boyhood of the men in question or because the data on their precocity simply failed to be recorded in their biographies. Nevertheless, the figures represented by the *uncorrected* estimated I.Q.'s alone are salient enough to give emphatic support to the author's conclusion.

176. Cuneo, Irene, and Terman, L. M. "Stanford Binet tests of 112 kindergarten children and 77 repeated tests." *Ped. Seminary,* 25: 1918, 414-428.

(See Dickson's tabled summary, No. 177.)

177. Dickson, V. *Mental Tests and the Classroom Teacher.* New York, 1923, 231 pp.

In pp. 65 ff. dealing with I.Q. constancy, original data are presented and other studies are also summarized.

Original Data. These data embrace 288 retests (or 576 individual tests) by 84 different examiners—chiefly by teachers who had been, or were being, trained in mental testing. The subjects ranged from 4 to 16 years, with a median I.Q. 86, and an I.Q. range from 33 to 140. The time between the tests varied from less than one month to three years, with a median interval of 13 months. Between the two tests the r was .90, the median difference 5.1 points, and the median increase .5 point.

The amount of change was distributed thus:

No. Points Difference	Percent of Tests
0 — 10	88.2
11 — 10	9.7
over 20	2.1

The probable causes of changes greater than 10 points were:

Foreign language difficulty.................17 cases
One or both tests incomplete............... 6 "
Psychopathic cases .... 3 "
No apparent cause.......................... 8 "

Summarized Data from Other Sources

Experimenter	r	Cases
Stenquist	.72	274
Rugg and Colloton	.84	137
Terman	.93	435
Baldwin	.74 to .94	(Various groups)
Gordon	.84	44
Bobertag	.95	...
Rosenow	.82	...
Cuneo and Terman	.85	31 (2-year interval)
Cuneo and Terman	.94	21 (6-month interval)
Cuneo and Terman	.95	25 (2-day interval)
Garrison	.88	298 (1-year interval)
Garrison	.91	127 (2-year interval)
Garrison	.83	42 (3-year interval)

178. Dvorak, August. "The relation of I.Q. to the prognosis of special-class pupils." *School and Soc.*, 19: 1924, 736-744.

A theoretical discussion explaining why low I.Q.'s should become still lower and high I.Q.'s should become still higher if the increments of yearly growth become consistently smaller with advancing C. A. If the rate of growth of bright and dull children with respect to the normal remains a constant, dull children will be slowing up in growth at a place on the M. A. scale where small increases of M. A. mean larger increases in absolute ability, and bright children will be tested at a place on the scale where small increases in absolute ability mean larger increases in M. A. Hypothetical growth curves are worked out upon this assumption and shown to be parallel to curves based on data from Kuhlmann's retests over 10 years of 639 feeble-minded children (*Jour. Appl. Psych.*, Sept., 1921).

The results of Woodrow (*Brightness and Dullness in Children*) and Baldwin and Stecher (*Mental Growth Curve of Normal and Superior Children*, 1922) bear out the author's contention.

179. Freeman, Frank N. *Mental Tests.* Cambridge, 1926, 503 pp.

On his p. 345, data from retests on the Stanford Binet are summarized as shown here on p. 323.

180. Garrison, C. "Additional retests by means of the Stanford Revision of the Binet-Simon Tests." *Jour. Educ. Psych.*, 13: 1922, 307-312.

(See tabled summaries by Dickson, No. 177, and by Freeman, No. 179.)

[To be read with Item 179, p. 322.]

Author	Number Cases	Percentage Differing 10 Points or More	Limits of Middle 50 Percent	Average Change	Coefficient of Correlation Between Two Tests
Terman	435	.15	−3.3 to +5.7	4.5	.93
Rugg and Colloton	137	.12	−2.3 to +5.6	4.7	.84
Garrison	468	.085	(−2 to +4) (−3 to +4) (−3 to +5)	5.4	.88
Rugg, L. S.	114		−1.2 to +1.9	3.1	.95

181. Gordon, Kate. "Some retests with the Stanford Binet Scale." *Jour. Educ. Psych.*, 13:1922, 363-365.

(See tabled summary by Dickson, No. 177.)

182. Gray, P. L., and Marsden, R. E. "The constancy of the intelligence quotient—final results." *Brit. Jour. Psych.*, 17:1926, 20-26.

Of a group of elementary pupils in a large city in England in 1920, 44 were tested twice, 12 three times, and 23 four times, at yearly intervals.

The Stanford Binet was used. All possible comparisons yielded 616 pairs of tests.

The central tendency of change is —1.3 points, and the P.E. of prediction is 5.5 points (based on the 616 pairs). The r for all these pairs is .85; for 100 pairs after first retest .88 ± .02; for 371 comparisons at last test, .85 ± .01.

183. Henmon, V. A. C., and Burns, Helen M. "The constancy of intelligence quotients with border-line and problem cases." *Jour. Educ. Psych.*, 14:1923, 247-250.

Fifty-nine cases who had an average I.Q. of 76 on the Stanford Binet gave a correlation of .91 between the first and second tests and an average difference of 5.3 points. The median loss on retest was 1.75 points. The median interval between tests was about two years.

184. Hildreth, Gertrude. "Stanford Binet retests of 441 school children." *Ped. Seminary*, 33:1926, 365-386.

"Data for this study were obtained from Stanford Binet test records of 441 Lincoln School pupils. The number of tests for each individual varied from two to eight, the ages from 3 to 18 years, I.Q.'s from 80 to 185 points. The tests were administered by 39 different examiners. . . . The median change in I.Q., when all possible pairs of

I.Q. were compared, was found to be .96; the limits of the middle 50 percent of variation, —3.50 and 5.71 points, Q, 4.61 points. The coefficient of correlation for all pairs was .814 ± .007." The median time-interval between tests was slightly over one year. Little difference in the constancy of retests was found for increasing time-intervals up to three years; but retests made after three years showed slightly greater discrepancies with the first tests.

185. Irwin, Elizabeth A., and Marks, L. A. *Fitting the School to the Child.* New York, 1924, 339 pp.

Normal children yielded I.Q.'s within 9 points of original I.Q.'s in 90 percent of 322 retests (p. 99), but only 53 percent of 30 neurotic children varied as little as 9 points.

186. Johnson, B. J. *Mental Growth of Children.* Dutton, N. Y., 1925, 160 pp.

Ten children first tested when three years old had an average gain in Stanford Binet I.Q. of 120 minus 112, or 8 points, with an average interval between tests of seven months. After an average interval of 14 more months, the average I.Q. was 123.

For 28 children first tested when four years old, the comparable figures were five points in 16 months, minus one point in 12 more months; for 15 first tested when five years old, plus five points in 12 months, minus one point in 11 more months. For 11 first tested when six years old, the same average was retained at the first and second retest (similar intervals).

The writers feel that their data indicate "greater inconstancy" of I.Q.'s in the early years.

Comment: Data on this point do not discriminate adequately between the possible 'instability' due to nursery-school training and that due to unreliability of the tests.

187. Kuhlmann, F. "The results of repeated mental reëxaminations of 639 feeble-minded over a period of ten years." *Jour. Appl. Psych.*, 5: 1921, 195-224.

The following average yearly decline in the I.Q.'s of children 7 to 20 was found for each of four grades of feeble-mindedness:

	Decline	Approximate I.Q. Range
Borderline	2.19 points	Over 74
Morons	1.21	50 to 74
Imbeciles	1.04	25 to 49
Idiots	.37	0 to 24

The I.Q.'s used in these computations were corrected for the faulty standardization of the scale.

Of the idiots 99 percent changed less than six points a year; of the imbeciles 98 percent changed less than six points; of the morons 98

percent changed less than nine points; of the borderline 94 percent changed less than 10 points.

The author shows the data upon decrease of I.Q. with age to be the natural consequence of an assumption of a decrease with age in the yearly increment of mental growth.

188. Poull, L. E. "Constancy of I.Q. in mental defectives, according to the Stanford Revision of Binet tests." *Jour. Educ. Psych.*, 12:1921, 323-324.

One hundred twenty-six defectives were tested and retested on the Stanford Binet at intervals of six months to three years. On the first test the ages ranged from 4 to 28 years, and the I.Q.'s from 20 to 90. The average increase on the second test was 1.28 points, and the middle 50 percent of changes extended from a decrease of 3.3 points to an increase of 4.8 points.

189. Rosonow, C. "The stability of the intelligence quotient." *Jour. of Delinquency*, 5:1920, 160-173.

(See tabled summary by Dickson, No. 177.)

190. Rugg, H., and Colloton, Cecile. "Constancy of the Stanford Binet I.Q. as shown by retests." *Jour. Educ. Psych.*, 12:1921, 315-322.

(See tabled summaries by Dickson, No. 177, and Freeman, No. 179.)

191. Rugg, L. S. "Retests and the constancy of the I.Q." *Jour. Educ. Psych.*, 16:1925, 341-343.

(See tabled summary by Freeman, No. 179.)

192. Terman, L. M. *The Intelligence of School Children.* Boston, 1919. 317 pp., esp. 138 ff.

Some 315 children retested in the vicinity of Stanford University yielded 435 comparisons on the Stanford Binet. The intervals between tests ranged from one day to seven years, and the age at the first test ranged from three to over 15 (8 cases). The correlation for the entire group of retests was .933. When differences in retest were tabled by age at first test and by interval between tests, the interval seemed to have no effect upon constancy. The age at the first test seemed to have little effect upon constancy at most ages, but the group tested first between three and six showed slightly more variability than other groups on the retests.

(See also the notes on Terman's data in the tabled summaries taken from Dickson, No. 177, and from Freeman, No. 179.)

9. Effect of Coaching

What are the limits of mental performance when training of a highly intensive and specialized nature is given?

193. *Bishop, O. "What is measured by intelligence tests?" *Jour. Educ. Research,* 9:1924, 29-38.

Sixty-four pupils from Grades VII to X were divided into the following groups: Groups A and B carefully matched from seventh and eighth grades on the basis of the first score on Otis Group Scale, Form A, chronological age, sex, apparent maturity, and habits of work; Groups C and D formed from ninth and tenth grades by similar pairings.

A series of 10 lessons similar to the 10 pages of the Otis Test was prepared by the author. The first five lessons were taught to Group A, and the last five lessons to Group B. All lessons were taught to Group C, and none to Group D. About 150 minutes was spent by Group A and B pupils and "more time" by Group C pupils.

After coaching, Group A showed "a total gain in scores of 376 points on the first five pages of the Otis scale as against 42 points on the last half." This is equal to a median percentage gain of 40 percent on the first half and 6 percent on the second half. Corresponding percentage figures for Group B (which had been trained on the second half) were 31 percent and 15 percent. Group C "made a total gain of 672 points, as against a total gain of 243 points" by Group D; *i.e.,* there was a median percentage gain of 30 percent for Group C against 11 percent for Group D.

Comment: It would have been desirable to report these gains in terms of some measure of dispersion instead of "percent of mean score" or total points gain.

194. *Chapman, A. E. "The effect of school training and special coaching on intelligence tests." *Forum of Educ.,* 2:1924, 172-183.

The following battery of tests was given in two forms to 300 elementary-school pupils between 11-0 and 12-11:

(1) Five arithmetic questions from Alpha
(2) Ten from Burt's reasoning
(3) Five from Alpha practical judgment
(4) Ten from Alpha number series
(5) Five from Alpha analogies
(6) Ten from Alpha disarranged sentences.

After the first composite test, coaching was given 15 minutes a day for two weeks, during which period "the teachers dealt with the questions in the test paper already set to their pupils, discussing with them the methods of solving them."

The S. D. on the first composite test for a single age group was 9 to 10 points. The average gain of the coached group was 9.4 and of the uncoached group was 2.3, leaving 7.1, or half a S. D., in favor of the coached group.

195. DeWeerdt, Esther H. "The transfer effect of practice in related functions upon a group intelligence test." *School and Soc.*, 25: 1927, 438-440.

Forty-five fifth-grade children of a test group and 45 children of a control group were given the Illinois Intelligence Examination twice, 12 days apart. The test group was given practice on 11 consecutive school days for an hour a day on the following features of intelligence tests: symbol-digit, addition, reading (Chapman-Cook speed test), cancellation, multiplication, same-opposite, multiplication by substitution.

The percentage gains upon the portions of the intelligence test after this training were as follows:

	Percent Gains	
	Test Group	Control Group
Analogies	115	68
Arithmetic problems	8	8
Sentence vocabulary	26	13
Substitution	30	31
Verbal ingenuity	13	16
Arithmetic ingenuity	12	15
Synonym-antonym	48	21
Average gain	36	25

The author concludes that analysis of results "indicates that the group intelligence test is a much more stable and reliable instrument for measuring group averages than its critics at times are willing to admit."

Comment: Since no measures of dispersion are published, the percentage gains are all the evidence we have. On the face of the figures, the advantage of the test group seems rather slight.

196. *Odell, C. W. "Some data as to the effect of previous testing upon intelligence scores." *Jour. Educ. Psych.*, 16: 1925, 482-486.

(1) Out of 5283 high-school seniors, those who had had an intelligence test before scored only two points higher on the Otis Self-Administering Test than those who had never had an intelligence test before.

(2) Army Alpha was administered to 44 pupils of ages 11 to 16, "after which several weeks of practice and coaching upon material

similar to that obtained in the test was given, accompanied by occasional retesting with the several forms of Alpha." Soon after the conclusion of the experiment ("a few days or even weeks later") Army Alpha was repeated and again "more than two years later." Because of uncertainty as to adult mental level, I.Q.'s were computed on three bases: (a) exact C. A.; (b) 14 years C. A., and (c) 16 years C. A., with the results here shown:

	Basis of Computing Adult I.Q.'s		
	Exact C.A.	16 Years	14 Years
Before training	119	119	120
Soon after training	154	154	157
Over 2 yrs. after	115	118	130

Thus, the median I.Q. was slightly less at the end of two years than at the beginning by the first two methods, and only 10 points greater when using the 14-year basal age. The increase shortly after training is 35 to 37 I.Q. points, depending upon the basal age employed.

Comment: The study shows unquestionable immediate effects of training. Effects after two years are not clear, since Alpha norms for unselected children in their teens are not available.

197. *White, W. "The influence of certain exercises in silent reading on scores on the Otis Group Intelligence Test." *Elem. School Jour.*, 23: 1923, 783-786. Also in *Ed. Admin. and Supervision*, 9: 1923, 179-182.

Various groups were given the Otis Test twice (at intervals of 36 to 59 days). During these intervals, reading drills, constructed to increase speed of reading, were given to all except 59 eighth-grade pupils who served as a control group. The following results were tabulated:

N	Group	Test 1	Test 2	Average Gain
85	4th, 5th, and 6th grade	72.3	80.9	8.6
43	8th grade	120.2	137.0	16.8
24	8th grade	99.4	122.6	23.2
60	7th grade	96.2	121.2	25.0
59	Control 8th grade	110.0	119.6	9.6

The author finds also, in further statistical treatment, a considerable inverse relation between percentage of increase and the Otis index of brightness, from which he concludes that abilities have been equalized through training.

Comment: It is unfortunate that no measures of dispersion are given, but the net gain of drilled groups seems to be six months to a year of normal growth.

The inverse relation between percentage increase and index of brightness might be due simply to the fact that in the duller groups the numerators from which the percentage of increase is calculated are smaller. Growth is not often found to be a multiplying factor times "present level" and could hardly be expected to be such a factor here.

II. Achievement

10. Family Resemblance

This topic is significant in the same way that family resemblance in *intelligence* is significant, but the data are no more crucial in this case than in the other. It is worth pointing out that here, as in the case of the intelligence studies, not a single one of the investigations we summarize reports correlations which have been corrected for attenuation. Hence, it should be borne in mind that the true resemblances are in practically every instance larger than the figures, based as they are upon scores from imperfect measuring instruments, would indicate.

198. Cobb, Margaret V. "Preliminary study of the inheritance of arithmetical abilities." *Jour. Ed. Psych.*, 8: 1917, 1-20.

Eight families including both parents and children over 14 (of which there were 20) were given the Courtis standard tests. The scores were measured in S. D. units derived from tests on 200 university undergraduate and graduate students.

Correlations between scores were:

	Child and Mid Parent	Child and Like Parent	Child and Unlike Parent
Addition	.28	.54	.05
Subtraction	.01	.27	−.11
Multiplication	.55	.79	.24
Division	.31	.44	.08
Copying figures	.45	.65	.13
Average	.32	.54	.08

Comment: The author attempts to interpret the correlations of the child's scores with those of the parent most like, and most unlike himself, upon the theory of Mendelian segregation. However, the assumption that graduated traits, such as arithmetic abilities, are due to

single Mendelian genes seems unwarranted, and hence the significance of the particular treatment in question is not clear.

199. Dexter, Emily S. "On family resemblance beyond the first degree of relation." *School and Soc.*, 19: 1924, 501-502.

Scholarship averages of 201 cousin pairs who had attended the University of Wisconsin correlated .21 ± .04; of 64 cousin pairs who had attended Agnes Scott College, .19 ± .08; of 96 pairs of uncles and aunts versus nephews and nieces at Wisconsin, .06 ± .07; of 18 pairs of aunts and nieces at Agnes Scott, .15 ± .16.

When tested by two or more tests out of Haggerty, Detroit, N. I. T., Stanford, or Herring-Binet, the average scores of 135 pairs of cousins in a Madison elementary school correlated .22 ± .06.

Comment: It is a questionable procedure to lump results from different types of tests in a single correlation, as was done in the last treatment. The method of obtaining average scores for this treatment is not explained.

200. Earle, E. L. "The inheritance of the ability to learn to spell." *Columbia University Contrib. to Phil., Psych. and Educ.*, 11: 1903, No. 2, 41-44.

An unstandardized spelling test was given to 191 pairs of siblings in a public school and 196 pairs in a parochial school. The author thinks the latter represents a more unselected group in the upper grades. The mark assigned to each child "was his distance above or below the average for his grade and sex in terms of the average deviation of that grade and sex as a unit."

Product-Moment Correlations

	r	N
Public school		
Brother-brother	.22	45
Sister-sister	.32	62
Brother-sister	.22	84
General	.25	191
Parochial school	r	N
Brother-brother	.56	48
Sister-sister	.48	44
Brother-sister	.50	104
General	.51	196

Comment: The study has historical significance, but the method of computing the scores makes precise interpretation impossible.

201. Gowen, J. W., and Gooch, M. "The mental attainments of college students in relation to the preparatory school and heredity." *Jour. Educ. Psych.*, 17: 1926, 408-418.

The following correlations between sibling pairs graduated from the University of Maine since 1921 or in college during 1922-1923 were based upon marks in college courses:

Subject	r	P.E.	N
English	.15	.04	279
Chemistry	.16	.05	159
Algebra	.16	.06	103
Analytical geometry	.22	.07	79
Advanced French	.19	.10	39
Elementary German	.41	.11	24

The average for brothers with brothers was .187 for the first four subjects (the only like-sex comparison with sufficient numbers for comparison). Since the average *r* for all sibs on the same four subjects is .171, it seems justifiable to lump the sexes.

202. Hildreth, Gertrude H. *The Resemblance of Siblings in Intelligence and Achievement.* New York, 1925, 65 pp. (Teachers College, Columbia University, Contributions to Education, No. 186.)

(See summary No. 11.)

203. Huestis, R. R., and Otto, T. P. "The grades of related students." *Jour. of Heredity,* 18: 1927, 225-226.

"Grade-point" scholarship averages for the sophomore year at the University of Oregon were secured for a hundred pairs of siblings. The correlations obtained were:

	r	P.E.	N
Sisters	.61	.07	38
Brothers	.74	.06	26
Brothers and sisters	.04	.11	36

A possible reason suggested to account for the higher values found for the like-sex coefficients than for the brother-sister coefficient is that "brothers are more liable than their sisters to be partially self-supporting and more liable to be interested in athletics."

Comment: The large size of the like-sex coefficients reënforces the author's suggestion that external conditions have affected the sibling similarity, since a correlation explicable by heredity alone would not have a value much above .50.

204. Peters, W. "Ueber Vererbung psychischer Fähigkeiten." *Fortschr. der Psych.,* 3: 1915, 185-382.

This study deals with the final sets of elementary-school marks of three generations of village and country people. Records were obtained for 1162 children, 344 pairs of parents, 177 grandparents, and 11 great-grandparents. The author attempted roughly to equate on a single scale

the marking systems of different localities. Marks were also equated for sex differences.

The following tabulation shows that, even when parents have the same marks, the children have better marks when grandparents are highest.

Marks of Parents	Marks of Grandparents	Marks of Children
1 – 1	1.25	1.19
	1.94	1.71
2 – 2	2.09	1.97
	2.70	2.23
3 – 3	2.38	2.00
	3.50	2.70

Taking the individual marks in four school subjects and constructing a fourfold table for mid-parent and offspring by referring each mark to the average of marks in all subjects, Yule's coefficient q was .455. Pearson's contingency coefficient .31, his mean-square contingency coefficient .36, and his fourfold r .37. Other q coefficients (based on all marks, but not on all marks *averaged* for each person), were:

Fathers and children	.34	
Mother and children	.48	
Mid-parent and son	.40	
Mid-parent and daughter	.51	
Father and son	.32	
Mother and son	.42	
Father and daughter	.36	
Mother and daughter	.55	
Child and average of 4 grandparents	.37	(151 children)
Siblings	.53	
Brothers	.58	
Sisters	.73	

Data are also given upon correlations between siblings on a few simple mental and motor tests. These are not summarized here.

Comment: The facts of this study have historical importance and establish unquestionably some family resemblance in school accomplishment. But the coefficients of family resemblance can be taken as little more than suggestive, because (1) school marks are always unreliable as individual measures, (2) the correction for varying standards of marking in different schools was only approximate, (3) Yule's coefficient generally gives values higher than a product-moment r.

205. Schuster, E., and Elderton, Ethel M. "The inheritance of ability." Eugenics Lab. Memoirs, 1: 1907, 1-42.

(1) Evidence of Oxford Class Lists. Number of fathers and sons and brothers who obtained first, second, third, and fourth-class honors, pass degree, and no degree were tabulated. All possible pairings were used.

Contingency coefficients for several thousand pairs of fathers and sons completing their work during the 19th century clustered around .30; for 4000 pairs of brothers, contingency coefficients were about the same, but four-fold correlations were approximately .40.

(2) Harrow and Charterhouse lists. "Position" in the school was found for pairs of brothers at equivalent times in their school careers in Harrow, and in Charterhouse. Between 1000 and 2000 pairs entered into various treatments.

Harrow pairs dated from 1858 "onwards."

C_1 varies from .37 to .44; mean .395
C_2 varies from .38 to .46; mean .415
r varies from .25 to .49; mean .398

206. Starch, D. "Inheritance of abilities in school studies." *School and Soc.*, 2: 1915, 602-610.

(1) In 63 families each having two or more children in a certain school (and yielding 111 sibling pairs when all possible pairs were used) the rank-order correlation for school marks during one year was .52.

(2) Standard tests given to 57 pairs of siblings in two elementary schools yielded the following correlations:

	r	N
Spelling	.21	57
Reading	.49	57
Writing (speed)	.18	24
Writing (quality)	.06	24

Average marks over a period of three to eight years computed for 38 elementary-school children of 11 families gave the following correlations for sibling pairs:

	r	N
Arithmetic	.32	54
Spelling	.21	54
Reading	.31	54
Language	.24	54

207. Starch, D. "Similarity of brothers and sisters in mental traits." *Psych. Rev.*, 24: 1917, 235-238.

Eighteen pairs of adult siblings, aged 19 to 32, at the University of Wisconsin were tested twice on two different days on a battery of achievement tests and simple mental tests. The average of the separate correlations for all the achievement tests (of which there were 10) was .42, of the mental tests (including memory, cancelling a's, can-

celling geometric forms, and tapping) .38. The correlation based on ranks in all tests combined was .73.

The author suggests as an argument for heredity that resemblance is apparently no greater in traits affected by school work than those not so affected.

Comment: It was not stated how the ranks in all tests were combined for the correlation .73.

Since the group used was undoubtedly highly selected, and since their variability is not reported, the actual values of the correlations can be taken only as suggestive.

11. Contribution of Intelligence to Achievement

The countless studies showing correlations between intelligence and achievement varying from close to zero for handwriting up to .70 or .80 for subjects like reading have not been summarized here because of limited space, and because they provide no means within their own data for determining whether the observed correlations are due to the influence of intelligence upon achievement or of achievement upon intelligence. The few studies contained in this section deal chiefly with the limitations upon achievement set by mental levels of various grades.

208. Burt, C. *Mental and Scholastic Tests.* London, 1921, 432 pp.

(See Summary, No. 125.)

209. Courtis, S. A. *Why Children Succeed.* Detroit, 1925, 271 pp.

A multiple correlation computed for measures of 165 boys between 10.5 and 13.5 upon the Stanford Achievement Test with two group intelligence tests, number of semesters of school attendance, and chronological age was .91. The standard error of estimate was 5.46.

Since this 5.46 is about 10 percent of the average Stanford Achievement score of the pupils, the author concludes that the factors under discussion "account for 90 percent of the changes in children." Since "the standard deviation of the changes in the achievement test scores on a retest after an interval of two days was a little under 7 percent of the mean score," he concludes "that all the remaining factors which influence a child's score have a combined general effect which is not more than 3 percent of the total score."

Comment: The statistical methods by which these conclusions are reached cannot be defended. The division of total score by standard error of estimate to secure "proportional contributions" is meaningless,

since (a) there is no inherent relationship between a mean and a standard error of estimate; (b) the mean of a test is an arbitrary distance from zero and can be raised or lowered at will, by inserting or deleting items easy enough to be passed by every pupil; (c) the sum of the separate standard errors of estimate for independent factors is inherently greater than the standard deviation of a criterion itself; (d) there is nothing in the data themselves to warrant the assumption of a causal influence of the factors in question upon achievement. The study contributes nothing beyond establishing certain correlations between variables.

210. Courtis, S. A. "The influence of certain social factors upon scores in the Stanford Achievement Tests." *Jour. Educ. Research,* 13:1926, 311-324, and 14:1926, 33-42.

Using data similar to those enumerated in the abstract No. 209 of *Why Children Succeed,* the author modified the statistical treatment originally applied and reached the conclusion that environmental factors account for 22 percent of the variation in school achievement of children 10.5 to 13.5 years old. This percentage value was arrived at by considering the standard error of estimate of the Stanford Achievement scores of 443 children after a multiple correlation of achievement with intelligence, age, and semesters spent at school had been computed. The standard error of estimate was found to be 44 percent of the total standard deviation of achievement scores; and since "chance variation" itself amounts to approximately 22 percent, 44 minus 22, or only 22 percent of the variation "is the maximum that can be predicted from environmental factors."

The author also attempted to appraise the effect of Sunday-School attendance upon the achievement level of the pupils and reached the conclusion that for children equal in intelligence, age, and school attendance, "The effect of Sunday-School attendance upon children coming from religious homes is to raise their Stanford scores approximately 5 percent (3.74 points) of the total value."

Comment: The effect upon achievement attributed to environment is said to be 22 percent in this study as compared with 3 percent in the previous study. But the modified treatment used here appears to introduce new errors in the underlying statistical theory. Standard deviations and errors of estimate cannot be added and subtracted to yield the type of interpretation sought by the author, because standard errors of estimate for various unique factors *do not combine* by simple addition to give the total standard deviation of a group.

The method by which the second conclusion is reached is not clear throughout, but some kind of environmental force appears to be at least a possibility. It is not possible to say whether this is religious or otherwise, since Sunday-School attendance may or may not be correlated with

other factors in environment which are actually responsible for the relationship found.

Commenting upon the small percent contributed by the Sunday-School, the author makes the statement, without clear foundation, that his results prove "not that the powers of religion are a negligible quantity, but that under present conditions it is only the stray individual who actually makes use of the rich gifts religion has to offer."

211. Dickson, V. E. "What first-grade children can do in school as related to what is shown by mental tests." *Jour. Educ. Research,* 2:1920, 475-480.

Twenty-five children with M. A.'s from three to six, and border-zone I.Q.'s, had had from two to four terms of previous schooling. After two and three-quarters terms under an especially good teacher in a special class, only six of the 25 could read in an easy primer, and these could read less well than the pupils of a normal class who had had only three-quarters of a term of instruction.

Of 29 children with slightly better I.Q.'s, and M. A.'s from four to six and who had had one to three terms of schooling, only seven learned to read approximately as well as the average high-first pupil after two and three-quarters terms of special work.

By way of contrast, only five out of 42 children testing normal or above failed at the end of the first grade; and these failed, according to their teacher, because of irregular attendance or excessive timidity.

Of 30 children who tested below 6 years M. A. and who had had one to two terms of previous schooling, only two were promoted to the high-first grade after two terms of special work.

Comment: The study shows clearly the limitations upon achievements set by intelligence, but does not provide a means for distinctly separating nature from nurture.

212. Franzen, R. H. *The Accomplishment Ratio.* New York, 1922, 59 pp. (Teachers College, Columbia University, Contributions to Education, No. 125.)

Of a group of 200 children serving as subjects for the author's experiment, 48 took the following battery of achievement tests four times during a period of two school years:

1. Woody-McCall Mixed Fundamentals (given only twice)
2. Thorndike Reading Scale Alpha
3. Thorndike Visual Vocabulary Scale, A2
4. Kelley-Trabue Completion Exercises in Language

They were also tested by the Stanford Binet at the beginning of the experiment.

Pupils were classified in each school subject on the basis of the achievement tests; their abilities in these subjects were pushed, and

promotions were made as often as necessary to maintain proper classification. Correlations of I.Q. with achievement rose from November, 1918, to June, 1920, as follows:

I.Q. with Vocabulary Quotient....................	.72 to .81
I.Q. with Reading Quotient......................	.64 to .79
I.Q. with Kelley-Trabue Completion Quotient.......	.63 to .84
I.Q. with Arithmetic Quotient (during second year experiment only)..........................	.46 to .73

The theoretical limit of these correlations (based as they are upon fallible tests) is about .95.

The author concludes that "intelligence is far and away the most important determinant of individual differences in product."

Comment: It is highly significant that better prediction of school achievement from intelligence could be obtained for results two years *after* the Binet tests than for results *at the time* of the Binet tests—under the condition that opportunity was "equalized" during the two years. This indicates that the Binet tests are probably measuring a factor in capacity to *do* school work which is to some degree independent of the *results* of school work.

One possibility not made clear in the study should be known before the results can be accepted without question. This concerns the relative effort that was made to push abilities before and after the school achievement of individual pupils became commensurate with their intelligence. If there were a difference between motivation before and after that point, correlations might have been due to that difference rather than to "equalized opportunity."

213. Merrill, Maud A. *On the Relation of Intelligence to Achievement in the Case of Mentally Retarded Children.* Baltimore, 1923-1925, 100 pp. (Comp. Psych. Monographs, 2: No. 10.)

Tests given to several hundred children, including subnormal, unselected, and gifted pupils, offered a basis for the following conclusion: "When retarded children are matched mental age for mental age by normal children of average I.Q., the normals attain a slightly higher average achievement score on a battery of educational tests, including tests of reading, writing, arithmetic, and spelling. The retarded pupils start a little ahead of the normals at M. A. seven, but fall behind as the comparisons proceed up to M. A. eleven, except on the writing test, on which they keep above the normals up to eleven and then fall only very slightly below."

At M. A. nine the retarded excel the gifted by about a year of E. A. They do not at M. A. eleven, however. At this level both the gifted and the retarded are from two to ninth months below this M. A. on educational subjects, except reading, on which the gifted are about five months in advance of their C. A.

12. Effect of Length of School Attendance

Investigations in this field are significant when neutral results are found; but when plus relationships are found, the possibility of selection cannot usually be satisfactorily measured and eliminated. Studies using controlled experimental methods over a term of years are needed.

214. *Corman, Oliver P. *Spelling in the Elementary School.* Boston, 1902, 98 pp.

When spelling was eliminated from the curriculum for several hundred Philadelphia school children during three years, it was found that the children did as well in spelling measured on a composition test, on lists of dictated words, and on a 15-minute test of spontaneously written words as the medians for the remaining school children of the city who had had the ordinary amount of drill. On a term examination paper required of all the city schools the undrilled children ended up four points below the city medians, but this difference was thought small as compared with fluctuations of 40 points in the medians of various classes of the same grade throughout the city.

Tabulating the time per day devoted to spelling in 11 schools (which varied from 10 to 40 minutes) it was found that "no relation between [time allotments] and the spelling percentages seems to obtain."

The conclusion is that "an influence, the suspension of whose operation for three years is not plainly manifested—is of so little importance as to be practically negligible."

Comment: Possibly, the slightly inferior performance of the experimental schools on the term examination may have been due to the type of the test, which was one set by the superintendent's office. It may have been based exclusively upon words actually studied in the text book.

215. Gordon, Hugh. *Mental and Scholastic Tests Among Retarded Children.* London Board of Education, Ed. Pamphlets, No. 44, 1923, 92 pp.

(See summary in section on schooling and intelligence, No. 127.)

216. Kelley, T. L. *The Influence of Nurture upon Native Differences.* New York, 1926, 49 pp.

The author wishes to find out whether unevenness in school abilities tend to increase or decrease under the influence of the schooling process. Scores upon the eight tests from the Stanford Achievement battery were used in comparisons made between eight-, eleven-, and fourteen-year-olds, all 'at age' for their grades. Assuming that the contribution of nature to the unevenness, or 'idiosyncrasy,' of an individual

upon any pair of school abilities remains constant throughout life, and that the influence of nurture upon idiosyncrasy is proportional to the rate of growth of school abilities, the author sets up equations for establishing the relative contributions of nature and nurture to adult idiosyncrasy.

He reaches the result that the only school abilities of the battery whose comparisons with the other abilities show an average increase of idiosyncrasy due to nurture are science information, history and literature information, language usage, and spelling. Nurture decreases the average "innate distinction" of paragraph meaning, sentence meaning, word meaning, computation, and arithmetic reasoning—each compared in turn with all of the other seven abilities.

Kelley's technique for determining idiosyncrasy is also applied to the Stanford Achievement scores of a group of eight-year-old gifted children. The results indicate that the idiosyncrasies for all the pairs of tests are in gifted children reduced during the three years between age eight and age eleven.

The conclusion is drawn, and its implications deplored, that schooling more frequently tends, in both normal and gifted groups, to level talents which are natively distinct than to accentuate idiosyncrasies.

Comment: The author admits that the study is based upon assumptions some of which are in need of corroboration. The following questions would apparently need to be answered before the conclusions could be considered as firmly established, particularly in the precise numerical form in which they are presented:

1. Would idiosyncrasies surely remain constant throughout life if it were not for the influence of nurture, or might some changes in unevenness occur that were due to nature or to different rates of development of different abilities?

2. Is the effect of nurture at any given time proportional to the rate of mental growth?

3. Are the proportions of children promoted or failed in school at different ages real functions of the relative variabilities of children of these ages in school abilities?

4. Are standard (z) scores adequate measures of children whose ages span an entire year if the norms used are those for children just six months away from a birthday—especially if the children are an age at which rapid growth is taking place, and more rapid growth in some abilities than in others?

217. Odell, C. W. "The effect of attendance upon school achievement." *Jour. Educ. Research,* 8: 1923, 422-432.

Several thousand school children from Grades I to VIII were given intelligence tests of various types and a battery of standard achievement tests, once at the beginning of a semester and again towards the close.

Their school marks and percentage attendance during the semester were also noted.

The correlation of attendance with increase in average achievement for 13,700 cases was .01; with average school mark (22,800 cases), .14; with intelligence quotients (3,800 cases), .06; and with attendance for the following semester (5,600 cases), .19.

Comment: The median attendance was reported as 94 percent. Less than 2 percent attended as little as 81 to 90 percent of the time, and only a fifth of 1 percent attended 10 percent of the time or less. Thus, the spread was so slight that differences in effect on achievement had little chance to show up.

218. *Rice, J. M. "The futility of the spelling grind." *Forum*, 23: 1897, 163-172, 409-419.

On a series of dictation tests and composition tests in spelling given to over 25,000 elementary-school children a table is published of average scores of classes in different cities with respect to time allotted to spelling. No r's were computed, but the author believes his data prove "that the results obtained by 40 or 50 minutes' daily instruction were not better than those obtained where not more than 10 or 15 minutes had been devoted to the subject."

Comment: By inspection, the conclusion drawn from these tables seems justified.

219. Snedden, D. "A note on the relationship between the number of months of study and proficiency (geometry)." *Jour. Educ. Psych.*, 17: 1926, 57.

Among applicants to Cooper Union night courses 441 boys were given Parts 1 and 2 of the Hawkes-Wood Placement Test in Plane Geometry. They noted on their test papers the number of months they had studied plane geometry.

The correlation of months studied with score was .22 (or .24 if corrected for the unreliability of the test). The mean number of months geometry was studied was 9, with an S. D. of 3.6.

Comment: No allowance was made for the amount of time elapsed since the boys studied geometry.

220. *Terman, L. M. *Genetic Studies of Genius, I.* Stanford University, 1925, 648 pp. (Pp. 304-305, "The influence of attendance upon educational accomplishment.")

The correlations between number of semesters of school attendance and achievement quotients for tests in the Stanford Achievement battery as computed for 109 gifted children between 10 and 11 years old, were as follows:

	r	P.E.
Spelling	.04	.07
Information	.00	.07
Reading	—.01	.07
Arithmetic	.13	.06
I.Q.	.01	.07

On the other hand I.Q. correlated with these subjects as follows:

	r	P.E.
Spelling	.33	.07
Information	.46	.05
Reading	.34	.06
Arithmetic	.26	.06

For this group, school attendance ranged from two to six and one-half years, with an S. D. of .86 years.

13. Effect of School Conditions, Size of Class, or Teaching Ability

There is little material available upon this topic, but what there is runs contrary to tradition. Several new studies published in this Yearbook deal more extensively with elements of this field than most of the studies summarized here.

221. Almack, J. C. *The Adaptation of the School Building to a Program of Educational Efficiency.* Stanford, Ph.D. dissertation, 1922. ("The relation between class size and efficiency in instruction—an experiment," pp. 31-41. Also in *Jour. N. E. A.*, 12: 1923, 107-109.)

A 5B class of 20 pupils was taught in two sections of 10 each during the winter quarter. These were united during spring quarter. A 5A class of 30 pupils was united during the winter, but divided into two groups during the spring. A 7B class had one section with 11 pupils and one section with 35, and both sections were kept constant during the two quarters. The Monroe Silent Reading tests were given to all pupils at the beginning and end of the winter quarter and at the end of the spring quarter. The Monroe Revised Diagnostic Arithmetic tests were given to the fifth, and the Woody-McCall test to the seventh grade. Student cadets were assigned to the different classes and told simply to "teach as much as possible."

A slightly larger percentage of children showed achievement gains in the large classes than in the small, but the difference was so small as to be ambiguous in the light of experimental conditions. The classes were not equated for initial achievement, excellence of teaching, etc. The author concludes: "Class size is not an important factor in determining efficiency in teaching as measured by standard tests."

222. *Breed, F. S., and McCarthy, G. D. "Size of class and efficiency of teaching." *School and Soc.*, 4: 1916, 965-971.

After 82 classes from Grades III to VII were drilled in spelling for 20 days on 80 words, large classes with an average enrollment of 45 surpassed small classes with an average enrollment of 27, but the percent improvement was only about 2 percent higher in the large classes than in the small classes.

Efficiency, in terms of percent improvement, showed a slight increase with increment in class size up to classes of 45-49, at which point efficiency slightly decreased (for classes of 45-49 and classes of 50-56, the largest classes used in the study).

223. Crabbs, Lelah M. *Measuring Efficiency in Supervision and Teaching.* New York, 1925, 98 pp. (Teachers College, Columbia University, Contrib. to Education, No. 175.)

As a criterion of teaching efficiency in 64 rural classrooms and 20 urban classrooms, the increase in accomplishment ratios of the 1100 pupils comprising the classes was taken. Merit rankings were also obtained upon the teachers from supervisory officials, and to the rural teachers two forms of the Steele-Herring Test of Professional Information (each involving two hours of testing time) were given.

The change in accomplishment ratio (on a combination of reading, arithmetic, spelling, composition, and penmanship) was measured for an interval of 12 months in the urban schools, and six months in the rural. The Herring-Binet or the National were the intelligence tests used.

The mean correlations of the changes in accomplishment ratio of five of the nine school subjects with the other four subjects were low, viz:

	Reading	Arith.	Spelling	Comp.	Penmanship
Urban	.26	.26	.14	.17	.28
Rural	.13	.10	.10	.20	.00

The author suggests that the low correlations between A. R.'s for different subjects may be due to the difference in emphasis of different teachers upon different subjects, yet calls the A. R. method "indisputably the most valid method yet devised for ascertaining the real efficiency of the teacher."

For rural teachers the correlation between A. R. measures and estimated efficiency was .33; for urban teachers —.25.

For the rural teachers the correlation between composite score of two Steele-Herring tests and A. R. measure was .05; but estimated efficiency with Steele-Herring correlated .41.

Practically no correlation was found between intelligence level or initial A. R.'s of the classes and their A. R. gains; so the author assumes that the A. R. measure of teaching efficiency need not make allowances for I.Q.'s or initial A. R.'s. The conclusion is drawn that

there is no relationship between what teachers know about educational ideas or between the supervisor's estimates of teaching ability and the objective A.R. measures of efficiency.

Comment: The extremely low correlations between A. R. gains on the five different subjects calls into question the validity of this measure of teaching efficiency. These low correlations, coupled with the fact that the A. R. compounds the errors of measurement both of intelligence and achievement batteries, and the fact that only six months to one year were allowed the teachers to bring about significant changes in their pupils, suggest that the author's conclusion is not thoroughly established.

224. Edmonson, J. B., and Mulder, F. J. "Size of class as a factor in university instruction." *Jour. Educ. Research,* 9:1924, 1-12.

Forty-four members of a large class in education (109 university students) and 43 members of a small class (45 students), who were equated for intelligence on the Army Alpha Test, averaged the same on the combined results of quizzes, mid-term examinations, and finals. The same instructor taught both groups and the same assistant graded their papers.

225. *Gowen, J. W., and Gooch, M. "The mental attainments of college students in relation to the preparatory school and heredity." *Jour. Educ. Psych.,* 17:1926, 408-418.

Results on nine hundred twenty-seven University of Maine graduates are reported. Only high schools represented by two or more University of Maine graduates between 1913 and 1921, inclusive, were considered. "Large high schools" had an average daily attendance of 100 or more, and "small high schools" less. Data are included, also, concerning pupils from private academies.

Following is a table of freshman average grades:

Schools	English		Chem.		Algebra		Anal.-Geom.		Advanced French		Elem. German	
	Mean	$P.E._m$	Mean	$P.E._m$	Mean	$P.E._m$	Mean	$P.E._m$	Mean	$P.E._m$	Mean	$P.E._m$
Large high . . .	82.6	0.3	80.1	0.4	85.4	0.5	80.9	0.5	88.9	0.5	83.4	0.6
Small high. . . .	82.0	0.7	82.9	0.7	82.4	1.1	83.8	1.0	. . . .		82.1	1.0
Academies. . . .	80.5	0.5	79.5	0.6	78.3	0.9	79.3	0.8	78.8	1.3	82.2	0.8

The authors conclude that "in general . . . it appears that the size of the high school has practically no relation to the average college marks which its pupils later receive."

226. *Stevenson, P. R. *Class Size in the Elementary School.* Ohio State University, 1925, 35 pp. (University Studies, 2: No. 10.)

In this experiment 26 pairs of second-grade classes, 20 pairs of fifth-grade classes and 16 pairs of seventh-grade classes were used. "Small classes" ranged in size from 16 to 32, "large classes" from 32 to 54. By shifting pupils into large or small classes at the end of the first school semester, comparisons were made on the basis of the same pupils, taught by the same teachers, when classes were large and when they were small.

All tests were reported in T scores derived from the groups tested in the experiment. As many large and small classes were tested during the first semester of the school year as during the second semester, so that neither large nor small classes would have the benefit of rapid initial improvement.

Advantages of large or small classes were reported as "percent that difference is of lowest gain" (during one semester). The results are summarized herewith in tabular form:

Subject	Size	Percent Advantage
Second Grade		
Arithmetic	Small	8.69
Language	Large	19.96
Reading speed	Large	19.06
Spelling	Small	1.81
All subjects	Large	2.53
Fifth Grade		
Language	Large	10.97
Reading speed	Large	249.78
Reading comprehension	Large	118.26
Arithmetic problems	Large	1.24
Analysis (problems)	Small	3.24
Spelling	Small	25.22
All subjects	Large	22.69

The author finds that "diminishing returns" begin when the class size reaches about 45 pupils.

The enormous saving that could be effected in large cities if class size could be increased by only one or two pupils is pointed out.

227. Thorndike, E. L., and Kruse, P. J. "The effect of humidification of a school room upon the intellectual progress of pupils." *School and Soc.*, 5: 1917, 657-660.

Two divisions of a sixth-grade of 43 pupils, the two being of equal initial ability, were so far as possible treated alike in every respect save the dryness of the school room. The experiment extended from December 4 to March 30, during which time the classroom of the "Dry" group received no humidification and averaged about 28.7 percent relative humidity, while that of the "Moist" group was kept at about 42.2 percent relative humidity.

On a battery of standard achievement tests given both at the beginning and end of the experiment, the superiority of one group to another was "no greater than its probable error. . . . What slight difference there is is in favor of the dry condition."

14. Effect of Emotional or Volitional Traits

As in the case of the corresponding section in the division on intelligence, few experimental data are available. Here appears to be a most promising field for future research.

228. Downey, June. "Testing the Will-Temperament Tests." *School and Soc.*, 16: 1922, 161-168.

Quoting a footnote, p. 163: "For a group of some seventy freshmen we find a correlation of the Thorndike score and average grade for two quarters of .40. A combined score on the Thorndike and total for Group Will-Temperament Test raised the correlation to .65."

229. Gilchrist, E. P. "The extent to which praise and reproof affect a pupil's work." *School and Soc.*, 4: 1916, 872-874.

The Courtis English Test 4B was given to 50 members of a class in educational psychology. The class was then divided into two groups of 25, placed in separate rooms, and the test repeated. One group was told that it did not do so well as the average 12-year-old; the other group was told it had done exceptionally well. The first group had scores totaling 59.80 and 59.21 on first and second test. The second group had 72.42 and 129.50, or an improvement of 79 percent.

Comment: The measure of improvement by percent is ambiguous.

230. Heck, W. H. "Correlation between amounts of home study and class marks." *School Rev.*, 25: 1916, 1-17.

Twenty-seven groups (divided by class, sex, number of units taken, etc., and therefore containing some of the same cases more than once) comprising over one thousand pupils in four Virginia high schools, showed correlations between number of hours of study during a representative week and average of school marks for the month in which that week occurred, of from $-.11 \pm .08$ to $.37 \pm .05$. The average of all the r's (for which no group has less than 57 students) is .084. There is enough variability in the hours of study and in the marks to make this zero relationship significant. The average minutes of study per week vary from about 350 to 750 with S. D.'s from 175 to 350 (approximately) in the different groups. The average marks vary from 81 to 87 with S. D.'s from 6 to 9 (approximately). Intelligence seems to overlay the effect of duration of study.

231. *Hurlock, Elizabeth B. "An evaluation of certain incentives used in school work." *Jour. Educ. Psych.*, 16: 1925, 145-159.

The subjects were 58 fourth-grade and 48 sixth-grade pupils in a Harrisburg school. They were evenly divided into four groups called the "control," "praised," "reproved," and "ignored" groups. The same sex distribution obtained in each group, and the same initial ability (within .03 point on the author's modification of the Courtis Research Test, addition form).

Fifteen minutes a day for five days were spent on similar tests, a different one each day, consisting of 30 examples of equal difficulty made up of six three-place numbers. The "control" group, after the first day, was separated from the others and took the remaining tests in another room. The "praised" group was called in front of the class for praise, the "reproved" group was each day reproved before the class for its performance, and the "ignored" group was simply ignored.

Initial scores for the four groups clustered about 11.8. S. D.'s for the groups varied from 4.2 to 7.7 on the various days.

Comparison of Final Tests

	Difference	P.E. Difference	Diff. ÷ P.E. Difference
Praised-control	8.9	1.1	7.8
Reproved-control	2.8	1.0	2.8
Ignored-control	1.0+	1.0−	1.1

232. Poffenberger, A. T., and Carpenter, F. L. "Character traits in school success." *Jour. Exper. Psych.*, 7: 1924, 67-74.

In a group of 97 elementary-school pupils, those who ranked higher on the National Intelligence Test than on school success showed a tendency to excel on certain elements of the Downey Will-Temperament Test.

Comment: The results are reported in such a way that the strength of the tendency is impossible to evaluate. Neither group dispersions of the Will-Temperament scores nor correlations of Will-Temperament with success are given.

233. May, Mark A. "Predicting academic success." *Jour. Educ. Psych.*, 14: 1923, 429-440.

With a group of 450 Liberal Arts College freshmen a combination of the Miller Test and Dartmouth Definitions correlated .60 with the number of honor points at the end of the first semester.

Honor points correlated with

Number of H. S. units offered at entrance	.22
High-school marks or entrance examination	.40
Hours of week spent in study	.32
Multiple of four factors above	.84

Leaving out hours of study, which correlated —.35 with intelligence, the multiple R was .64. Intelligence with hours of study constant correlated .80; with high-school average constant, .53; and with units constant, .59.

The conclusion is that a combination of intelligence and hours of study is the most reliable means of prediction. Influence of other factors is considered proportional to their partials with honor points when the remaining factors are partialled out.

Comment: The conclusion seems sound in the main, but the interpretations of the partial correlations are subject to some of the difficulties pointed out in Part I, Chapter II on "Statistical Hazards."

234. Gates, A. I. "Nature and educational significance of physical status and of mental, physiological, social, and emotional maturity." *Jour. Educ. Psych.*, 15: 1924, 329-358.

In the Horace Mann School 57 fourth-grade pupils were tested by the Stanford Binet and a battery of achievement tests, and rated on emotional maturity and social maturity by five judges. The reliability of the combined ratings was about .87. Correlations found were:

M. A. with social maturity	.20
M. A. with emotional maturity	.16
E. A. with social maturity	.19
E. A. with emotional maturity	.20

Comment: What slight correlations actually obtain here between qualities of personality and achievement may be due to causative interdependence or to "halo-effects" in the teachers' ratings.

15. Effect of Segregation According to Ability

The lack of very pronounced effects suggests that the advantages of segregation may be more in the development of habits of work, ideals, and interests than in the mere advancement of measurable school abilities.

235. Burtt, H. E., Chassell, L. M., and Hatch, E. M. "Efficiency of instruction in unselected sections in elementary psychology compared with that in sections selected on basis of intelligence tests." *Jour. Educ. Psych.*, 14: 1923, 154-161.

Three experimental classes having from 35 to 60 students were selected for high, medium, and low intelligence, respectively; six control classes taught by the same instructors were not sectioned for ability. The experimental classes had average percentile scores on a modified Army Alpha test of 83, 54, and 21 the first semester, and 84, 54, and 22 the second semester (based on norms for 2,000 freshmen).

During the first semester the rate of progress in all sections was kept normal, but in the second semester the top and medium sections and their two controls were 'pushed,' upon the principle that the correlation between intelligence and course marks would measure efficiency of instruction. Product-moment r's were computed and corrected for range by Kelley's formula. In both the high and medium sections, the correlations between intelligence and marks rose during the second semester, but they did not in the control groups. The conclusion was that segregation was not an advantage unless methods were then used especially adapted to ability.

236. Cobb, Margaret V., and Taylor, Grace A. "Stanford Achievement Tests with a group of gifted children." *Twenty-Third Yearbook* of this Society. Part I, 1924, 275-289.

The average accomplishment ratio on the Stanford Achievement Test of 52 children with I.Q.'s of 135 to 190 in "opportunity classes" at Manhattan P. S. 165 was 97 after one year of segregation. This accomplishment ratio is higher than that usually found for unselected gifted children.

237. Cook, R. R. "A study of the results of homogeneous grouping of abilities in high-school classes." *Twenty-Third Yearbook* of this Society. Part I, 1924, 302-312.

Pupils in the Topeka High School were divided, on the basis of previous term's marks in similar subjects or on the basis of Terman Group Test scores, into classes containing approximately the best half and poorest half of students. Mixed classes were also used for comparison. First-term freshman English, first-term sophomore English, plane geometry, and ancient history were dealt with. Two sets of high, low, and mixed classes were included for the different subjects—each set taught by a different teacher and the high and low sets "equated" for time of day. No attempt was made to adapt the subject matter to the types of classes taught.

Scores of averages are not reported. Neither are the numbers of pupils in each type of class. These facts would make it seem possible to accept only provisionally the author's conclusions, which are:

"Pupils of superior ability in geometry did better work in mixed classes than they did in selected groups. Their history scores were much higher than scores of pupils of the same ability in mixed classes. In English the differences were so small as to be negligible."

"Neither can we say that separate grouping is always beneficial to slow pupils. In this experiment slow pupils in separate groups obtained much better results in geometry, slightly better in English; but in history they were much retarded by being grouped separately." These conclusions are based upon tests devised by the teachers and administered to all the classes three or four times during the semester.

238. Laird, D. A. "A study of the influence of sectioning students upon their achievement." *Jour. Educ. Psych.*, 14: 1923, 143-153.

Sections in elementary psychology (1) between 1909 and 1915, and (2) between 1915 and 1921 (omitting 1919-20), involving 2700 students, were studied.

During the first period the sections met three times a week for lectures or experimental work. During the second period they met twice a week as a class and one hour in smaller conferences. Marks were from A to F, with C average. During both periods there was a nine o'clock section and an eleven o'clock section (whether taught by the same instructor is not stated).

Eleven o'clock sections during the first period had a smaller proportion of good and poor students. The first semester marks averaged 2.7 percent A's and 0.9 percent F's; the second semester marks averaged 3.4 percent A's and 0.9 percent F's. Corresponding figures for the nine o'clock section were 5.2 and 1.8; 6.2 and 0.4. Average grades were apparently about the same (but not reported) for the nine and eleven o'clock sections. No significant differences in grades were found during the second period (1915-1921), and differences during the *first* period are less during the second semester. For the first semester of 1921-22, Thorndike scores correlated with grades .21 for the nine o'clock section and .32 for the eleven o'clock section.

239. Moyer, E. L. "A study of the effects of classification by intelligence tests." *Twenty-Third Yearbook* of this Society, Part I, 1924, 313-322.

Freshmen pupils were segregated for algebra and Latin classes into three sections on the basis of two forms of the National Intelligence Test. Owing to conflicts in programs, not all could go into these segregated classes, which resulted in "mixed" classes for comparison. No attempt was made to differentiate content of instruction.

The reliability or exact significance of the reported differences is unknown, because numbers and dispersions in scores are not reported. The author draws the following conclusions:

(1) In algebra tests (Hotz Algebra Scale) the superior segregated pupils are about equal to the superior pupils in mixed classes, but in

Latin tests (Henmon Tests) an advantage appears in favor of superior pupils in segregated classes.

(2) In both subjects the medium pupils in superior or in medium segregated classes excelled those in mixed classes.

(3) In algebra the medium pupils in low segregated classes were lower than those in mixed sections.

(4) In algebra the poor pupils, both in low and mixed classes, made about the same scores.

Comment: The conclusions seem to have qualitative validity, even if they have quantitative ambiguity.

INFORMATION CONCERNING THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

1. **Purpose.** The purpose of the National Society is to promote the investigation and discussion of educational questions. To this end it holds an annual meeting and publishes a series of yearbooks.

2. **Eligibility to Membership.** Any person who is interested in receiving its publications may become a member by sending to the Secretary-Treasurer information concerning name, address, and class of membership desired (see Item 4) and a check for $3.00 or $2.50 (see Items 4 and 5). Membership may not be had by libraries or by institutions.

3. **Period of Membership.** Applicants for membership may not date their entrance back of the current calendar year, and all memberships terminate automatically on December 31st, unless the dues for the ensuing year are paid as indicated in Item 6. Unless special arrangements are made, applications received between January 1st and November 30th are considered as applications for membership during the current year, but those received during December are considered as applications for the year beginning the following January.

4. **Classes of Members.** Application may be made for either active or associate membership. Active members pay two dollars dues annually, receive two copies of each publication, are entitled to vote, to participate in discussion, and (under certain conditions) to hold office. Associate members pay dues of $1.50 annually, receive one copy of each publication, may attend the meetings of the Society, but may not vote, hold office or participate in discussion. The names of active members only are printed in the yearbook. There were in 1927 about 1250 active and 1300 associate members.

5. **Entrance Fee.** New active and new associate members are required the first year to pay, in addition to the dues, an entrance fee of one dollar.

6. **Payment of Dues.** Statements of dues are rendered in October or November for the following calendar year. By vote of the Society at the 1919 meeting, "any member so notified whose dues remain unpaid on January 1st, thereby loses his membership and can be reinstated only by paying the entrance fee of one dollar required of new members."

School warrants and vouchers from institutions must be accompanied by definite information concerning the name and address and class of membership of the person for whom membership fee is being paid.

7. **Distribution of Yearbooks to Members.** The yearbooks, ready prior to each February meeting, will be mailed from the office of the publishers, only to members whose dues for that year have been paid. Members who desire yearbooks prior to the current year must purchase them directly from the publishers (see Item 8).

8. **Commercial Sales.** The distribution of all yearbooks prior to the current year, and also those of the current year not regularly mailed to members in exchange for their dues, is in the hands of the publishers, not of the secretary. For such commercial sales, communicate directly with the Public School Publishing Company, Bloomington, Illinois, who will gladly send a price list covering all the publications of this Society and of its predecessor, the National Herbart Society. Bound sets can be supplied of the first twenty-six Yearbooks.

9. **Yearbooks.** The yearbooks are issued in parts (usually two) from one to four months before the February meeting. They comprise from

500 to 800 pages annually. Unusual effort has been made to make them, on the one hand, of immediate practical value, and on the other hand, representative of sound scholarship and scientific investigation. Many of them are the fruit of coöperative work by committees of the Society.

10. **Meetings.** The annual meetings, at which the yearbooks are discussed, are held in February at the same time and place as the meeting of the Department of Superintendence of the National Education Association.

Applications for membership will be handled promptly at any time on receipt of name and address, together with check for the appropriate amount ($3.00 for new active membership, $2.50 for new associate membership). New members will receive, about three weeks after their application is mailed, the yearbook slated for discussion at the February meeting of the year in which the application is made (except in the case of December applications, as explained in Item 3).

GUY M. WHIPPLE, Secretary-Treasurer.

10 Putnam Street,
Danvers, Mass.

PUBLICATIONS OF THE NATIONAL HERBART SOCIETY

(Now the National Society for the Study of Education)

	Postpaid Price
First Yearbook, 1895	$0.79
First Supplement to First Yearbook	.28
Second Supplement to First Yearbook	.27
Second Yearbook, 1896	.85
Supplement to Second Yearbook	.27
Third Yearbook, 1897	.85
Note—Ethical Principles Underlying Education, John Dewey. Reprinted from Third Yearbook	.27
Supplement to Third Yearbook	.27
Fourth Yearbook, 1898	.79
Supplement to Fourth Yearbook	.28
Fifth Yearbook, 1899	.79
Supplement to Fifth Yearbook	.54

PUBLICATIONS OF THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

	Postpaid Price
First Yearbook, 1902, Part I—Some Principles in the Teaching of History. Lucy M. Salmon	$0.54
First Yearbook, 1902, Part II—The Progress of Geography in the Schools. W. M. Davis and H. M. Wilson	.53
Second Yearbook, 1903, Part I—The Course of Study in History in the Common School. Isabel Lawrence, C. A. McMurry, Frank McMurry, E. C. Page, and E. J. Rice	.53
Second Yearbook, 1903, Part II—The Relation of Theory to Practice in Education. M. J. Holmes, J. A. Keith, and Levi Seeley	.53
Third Yearbook, 1904, Part I—The Relation of Theory to Practice in the Education of Teachers. John Dewey, Sarah C. Brooks, F. M. McMurry, *et al.*	.53
Third Yearbook, 1904, Part II—Nature Study. W. S. Jackman	.85
Fourth Yearbook, 1905, Part I—The Education and Training of Secondary Teachers. E. C. Elliott, E. G. Dexter, M. J. Holmes, *et al.*	.85
Fourth Yearbook, 1905, Part II—The Place of Vocational Subjects in the High-School Curriculum. J. S. Brown, G. B. Morrison, and Ellen H. Richards	.53
Fifth Yearbook, 1906, Part I—On the Teaching of English in Elementary and High Schools. G. P. Brown and Emerson Davis	.53
Fifth Yearbook, 1906, Part II—The Certification of Teachers. E. P. Cubberley	.64
Sixth Yearbook, 1907, Part I—Vocational Studies for College Entrance. C. A. Herrick, H. W. Holmes, T. de Laguna, V. Prettyman, and W. J. S. Bryan	.70
Sixth Yearbook, 1907, Part II—The Kindergarten and its Relation to Elementary Education. Ada Van Stone Harris, E. A. Kirkpatrick, Maria Kraus-Boelté, Patty S. Hill, Harriette M. Mills, and Nina Vandewalker	.70
Seventh Yearbook, 1908, Part I—The Relation of Superintendents and Principals to the Training and Professional Improvement of their Teachers. Charles D. Lowry	.78
Seventh Yearbook, 1908, Part II—The Co-ordination of the Kindergarten and the Elementary School. B. J. Gregory, Jennie B. Merrill, Bertha Payne, Margaret Giddings	.78
Eighth Yearbook, 1909, Parts I and II—Education with Reference to Sex. C. R. Henderson and Helen C. Putnam. Both parts	1.60
Ninth Yearbook, 1910, Part I—Health and Education. T. D. Wood	.85
Ninth Yearbook, 1910, Part II—The Nurse in Education. T. D. Wood, *et al.*	.78
Tenth Yearbook, 1911, Part I—The City School as a Community Center. H. C. Leipziger, Sarah E. Hyre, R. D. Warden, C. Ward Crampton, E. W. Stitt, E. J. Ward, Mrs. E. C. Grice, C. A. Perry	.78
Tenth Yearbook, 1911, Part II—The Rural School as a Community Center. B. H. Crocheron, Jessie Field, F. W. Howe, E. C. Bishop, A. B. Graham, O. J. Kern, M. T. Scudder, B. M. Davis	.79
Price for Yearbooks VI to X inclusive, 8 vo. cloth	5.00
Eleventh Yearbook, 1912, Part I—Industrial Education: Typical Experiments Described and Interpreted. J. F. Barker, M. Bloomfield, B. W. Johnson, P. Johnston, L. M. Leavitt, G. A. Mirick, M. W. Murray, C. F. Perry, A. L. Safford, and H. B. Wilson	.85
Eleventh Yearbook, 1912, Part II—Agricultural Education in Secondary Schools. A. C. Monahan, R. W. Stimson, D. J. Crosby, W. H. French, H. F. Button, F. R. Crane, W. R. Hart, G. F. Warren	.85

Twelfth Yearbook, 1913, Part I—The Supervision of City Schools. Franklin Bobbitt, J. W. Hall, J. D. Wolcott.......................... .85

Twelfth Yearbook, 1913, Part II—The Supervision of Rural Schools. A. C. Monahan, L. J. Hanifan, J. E. Warren, Wallace Lund, U. J. Hoffman, A. S. Cook, E. M. Rapp, Jackson Davis, J. D. Wolcott.......................... .85

Thirteenth Yearbook, 1914, Part I—Some Aspects of High-School Instruction and Administration. H. C. Morrison, E. R. Breslich, W. A. Jessup, and L. D. Coffman .... .85

Thirteenth Yearbook, 1914, Part II—Plans for Organizing School Surveys, with a Summary of Typical School Surveys. Charles H. Judd and Henry L. Smith .79

Fourteenth Yearbook, 1915, Part I—Minimum Essentials in Elementary School Subjects—Standards and Current Practices. H. B. Wilson, H. W. Holmes, F. E. Thompson, R. G. Jones, S. A. Courtis, W. S. Gray, F. N. Freeman, H. C. Pryor, J. F. Hosic, W. A. Jessup, W. C. Bagley.......................... .85

Fourteenth Yearbook, 1915, Part II—Methods for Measuring Teachers' Efficiency. Arthur C. Boyce .... .79

Fifteenth Yearbook, 1916, Part I—Standards and Tests for the Measurement of the Efficiency of Schools and School Systems. G. D. Strayer, Bird T. Baldwin, B. R. Buckingham, M. R. Trabue, F. W. Ballou, D. C. Bliss, H. G. Childs, S. A. Courtis, E. P. Cubberley, C. H. Judd, George Melcher, E. E. Oberholtzer, J. B. Sears, Daniel Starch, and G. M. Whipple.......................... .85

Fifteenth Yearbook, 1916, Part II—The Relationship Between Persistence in School and Home Conditions. Charles E. Holley.......................... .87

Fifteenth Yearbook, 1916, Part III—The Junior High School. Aubrey A. Douglass .... .85

Sixteenth Yearbook, 1917, Part I—Second Report of The Committee on Minimum Essentials in Elementary-School Subjects. W. C. Bagley, W. W. Charters, F. N. Freeman, W. S. Gray, Ernest Horn, J. H. Hoskinson, W. S. Monroe, C. F. Munson, H. C. Pryor, L. W. Rapeer, G. M. Wilson, and H. B. Wilson 1.00

Sixteenth Yearbook 1917, Part II—The Efficiency of College Students as Conditioned by Age at Entrance and Size of High School. B. F. Pittenger...... .85

Seventeenth Yearbook, 1918, Part I—Third Report of the Committee on Economy of Time in Education. W. C. Bagley, B. B. Bassett, M. E. Branom, Alice Camerer, J. E. Dealey, C. A. Ellwood, E. B. Greene, A. B. Hart, J. F. Hosic, E. T. Housh, W. H. Mace, L. R. Marston, H. C. McKown, H. E. Mitchell, W. C. Reavis, D. Snedden, and H. B. Wilson.......................... .85

Seventeenth Yearbook, 1918, Part II—The Measurement of Educational Products. E. J. Ashbaugh, W. A. Averill, L. P. Ayres, F. W. Ballou, Edna Bryner, B. R. Buckingham, S. A. Courtis, M. E. Haggerty, C. H. Judd, George Melcher, W. S. Monroe, E. A. Nifenecker, and E. L. Thorndike.......... 1.00

Eighteenth Yearbook, 1919, Part I—The Professional Preparation of High-School Teachers. G. N. Cade, S. S. Colvin, Charles Fordyce, H. H. Foster, T. W. Gosling, W. S. Gray, L. V. Koos, A. R. Mead, H. L. Miller, F. C. Whitcomb, and Clifford Woody 1.65

Eighteenth Yearbook, 1919, Part II—Fourth Report of Committee on Economy of Time in Education. F. C. Ayer, F. N. Freeman, W. S. Gray, Ernest Horn, W. S. Monroe, and C. E. Seashore.......................... 1.10

Nineteenth Yearbook, 1920, Part I—New Materials of Instruction. Prepared by the Society's Committee on the Materials of Instruction.......................... 1.10

Nineteenth Yearbook, 1920, Part II—Classroom Problems in the Education of Gifted Children. T. S. Henry.......................... 1.00

Twentieth Yearbook, 1921, Part I—New Materials of Instruction. Second Report by the Society's Committee.......................... 1.30

Twentieth Yearbook, 1921, Part II—Report of the Society's Committee on Silent Reading. M. A. Burgess, S. A. Courtis, C. E. Germane, W. S. Gray, H. A. Greene, Reginia R. Heller, J. H. Hoover, J. A. O'Brien, J. L. Packer, Daniel Starch, W. W. Theisen, G. A. Yoakum, and representatives of other school systems 1.10

Twenty-First Yearbook, 1922, Parts I and II—Intelligence Tests and Their Use. Part I—The Nature, History, and General Principles of Intelligence Testing. E. L. Thorndike, S. S. Colvin, Harold Rugg, G. M. Whipple. Part II—The Administrative Use of Intelligence Tests. H. W. Holmes, W. K. Layton, Helen Davis, Agnes L. Rogers, Rudolph Pintner, M. R. Trabue, W. S. Miller, Bessie L. Gambrill, and others. The two parts are bound together........ 1.60

Twenty-Second Yearbook, 1923, Part I—English Composition: Its Aims, Methods, and Measurement. Earl Hudelson.......................... 1.10

Twenty-Second Yearbook, 1923, Part II—The Social Studies in the Elementary and Secondary School. A. S. Barr, J. J. Coss, Henry Harap, R. W. Hatch, H. C. Hill, Ernest Horn, C. H. Judd, L. C. Marshall, F. M. McMurry, Earle Rugg, H. O. Rugg, Emma Schweppe, Mabel Snedaker, and C. W. Washburne 1.50

Twenty-Third Yearbook, 1924, Part I—The Education of Gifted Children. Report of the Society's Committee, Guy M. Whipple, Chairman............ 1.75

Twenty-Third Yearbook, 1924, Part II—Vocational Guidance and Vocational Education for Industries. A. H. Edgerton and others.......................... 1.75

Twenty-Fourth Yearbook, 1925, Part I—Report of the National Committee on Reading. W. S. Gray, Chairman; F. W. Ballou, Rose L. Hardy, Ernest Horn, Frances Jenkins, S. A. Leonard, Estaline Wilson, Laura Zirbes..... 1.50

Twenty-Fourth Yearbook, 1925, Part II—Adapting the Schools to Individual Differences. Report of the Society's Committee. Carleton W. Washburne, Chairman 1.50

Twenty-Fifth Yearbook, 1926, Part I—The Present Status of Safety Education. Report of the Society's Committee. Guy M. Whipple, Chairman.......... 1.75

Twenty-Fifth Yearbook, 1926, Part II—Extra-Curricular Activities. Report of the Society's Committee. Leonard V. Koos, Chairman.................. 1.50

Twenty-Sixth Yearbook, 1927, Part I—Curriculum Making: Past and Present. Report of the Society's Committee. Harold O. Rugg, Chairman........... 1.75

Twenty-Sixth Yearbook, 1927, Part II—The Foundations of Curriculum Making. Prepared by individual members of the Society's Committee. Harold O. Rugg, Chairman 1.50

Twenty-Seventh Yearbook, 1928, Part I—Nature and Nurture: Their Influence upon Intelligence. Prepared by the Society's Committee. Lewis M. Terman, Chairman 1.75

Twenty-Seventh Yearbook, 1928, Part II—Nature and Nurture: Their Influence upon Achievement. Prepared by the Society's Committee. Lewis M. Terman, Chairman 1.75

PUBLIC SCHOOL PUBLISHING COMPANY, BLOOMINGTON, ILLINOIS

Agents—The Baker and Taylor Company New York

Form No. 4.

BOOK CARD

Coll. No.............. Accn. No......

Author..................................

Title.....................................

Date.	Issued to	Returned on
....	....	....